GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: June 3, 2003, 14:05:55 ; Search time 51.4323 Seconds
(without alignments)
984.502 Million cell updates/sec

Title: US-09-804-690-4
Perfect score: 2002
Sequence: 1 MVSKYKYRDLTVRETVNVIT.....FQLRALMQKARKTAGLSDLY 380

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2002	100.0	380	18	AAW19111	Human tumour susce
2	2002	100.0	390	20	AAW93424	Human NHTS protein
3	1900.5	94.9	381	18	AAW19110	Mouse tumour susce
4	1900.5	94.9	381	20	AAW93425	Mouse tsg101 prote
5	1404	70.1	307	23	ABP41729	Human ovarian anti
6	1117	55.8	237	22	AAE09328	Human intracellula
7	809.5	40.4	331	22	ABB64607	Drosophila melanog
8	455	22.7	398	21	AAG06370	Arabidopsis thalia
9	455	22.7	412	21	AAG06369	Arabidopsis thalia
10	434	21.7	90	21	AAG01689	Human secreted pro
11	424	21.2	340	21	AAG06371	Arabidopsis thalia
12	380	19.0	322	21	AAG13898	Arabidopsis thalia
13	371	18.5	341	22	AAB93473	Human protein sequ
14	346	17.3	146	21	AAB53717	Human colon cancer
15	331	16.5	452	22	AAU23116	Novel human enzyme
16	267.5	13.4	288	21	AAG13899	Arabidopsis thalia
17	161.5	8.1	168	21	AAG13900	Arabidopsis thalia
18	161	8.0	485	21	AAB57086	Human prostate can
19	153	7.6	218	22	ABG18833	Novel human diagno
20	151	7.5	466	11	AAR07084	Recombinant human
21	148.5	7.4	148	21	AAY86515	Human gene 71-enco
22	148.5	7.4	148	23	ABB97431	Novel human protei
23	148.5	7.4	176	22	AAG75132	Human colon cancer
24	148.5	7.4	397	22	AAM39852	Human polypeptide
25	148.5	7.4	397	22	AAG81381	Human AFP protein
26	148.5	7.4	442	22	AAM41638	Human polypeptide
27	144.5	7.2	777	23	ABP41839	Human ovarian anti
28	142	7.1	505	20	AAY07117	Lung cancer associ
29	142	7.1	905	18	AAW31186	Human p160 polypep
30	142	7.1	1135	18	AAW31185	Human p160 polypep
31	141	7.0	244	21	AAG40015	Arabidopsis thalia
32	141	7.0	245	21	AAG05889	Arabidopsis thalia
33	141	7.0	245	21	AAG38059	Arabidopsis thalia
34	141	7.0	247	21	AAG05888	Arabidopsis thalia
35	141	7.0	247	21	AAG38058	Arabidopsis thalia
36	141	7.0	263	21	AAG38057	Arabidopsis thalia
37	141	7.0	274	21	AAG05887	Arabidopsis thalia
38	139.5	7.0	258	21	AAB44500	Plant viral moveme
39	138.5	6.9	827	23	ABP35602	Fungal ZBC protein
40	135.5	6.8	177	22	ABG09342	Novel human diagno
41	135.5	6.8	1277	22	ABB66688	Drosophila melanog
42	135	6.7	121	21	AAG21511	Arabidopsis thalia
43	134.5	6.7	172	22	ABG18832	Novel human diagno
44	133	6.6	144	22	AAM42061	Human polypeptide
45	132	6.6	371	19	AAW64535	Human leukocyte ce

ALIGNMENTS

RESULT 1
AAW19111
ID AAW19111 standard; Protein; 380 AA.
XX
AC AAW19111;
XX
DT 10-DEC-1997 (first entry)
XX
DE Human tumour susceptibility protein TSG101.
XX
KW Tumour susceptibility protein; TSG101; tumorigenesis;
KW breast cancer; diagnosis; therapy; human.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 130..205
FT /note= "proline-rich domain"
FT Domain 231..302
FT /note= "coiled-coil domain"
FT Modified-site 11
FT /note= "potential protein kinase C phosphorylation
FT site"
FT Modified-site 38
FT /note= "potential protein kinase C phosphorylation
FT site"
FT Modified-site 86
FT /note= "potential protein kinase C phosphorylation
FT site"
FT Modified-site 89
FT /note= "potential protein kinase C phosphorylation
FT site"
FT Modified-site 215

```
FT                  /note= "potential protein kinase C phosphorylation
FT                   site"
FT   Modified-site  225
FT                  /note= "potential protein kinase C phosphorylation
FT                   site"
FT   Modified-site  357
FT                  /note= "potential protein kinase C phosphorylation
FT                   site"
FT   Modified-site  38
FT                  /note= "potential casein kinase II phosphorylation
FT                   site"
FT   Modified-site  210
FT                  /note= "potential casein kinase II phosphorylation
FT                   site"
FT   Modified-site  249
FT                  /note= "potential casein kinase II phosphorylation
FT                   site"
FT   Modified-site  265
FT                  /note= "potential casein kinase II phosphorylation
FT                   site"
FT   Modified-site  290
FT                  /note= "potential casein kinase II phosphorylation
FT                   site"
FT   Modified-site  44
FT                  /note= "potential N-glycosylation site"
FT   Modified-site  150
FT                  /note= "potential N-glycosylation site"
FT   Modified-site  297
FT                  /note= "potential N-glycosylation site"
XX
PN   WO9718333-A1.
XX
PD   22-MAY-1997.
XX
PF   15-NOV-1996;   96WO-US18828.
XX
PR   13-JUN-1996;   96US-0670274.
PR   16-NOV-1995;   95US-0006856.
PR   16-JAN-1996;   96US-0585758.
XX
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
XX
PI   Cohen SN,  Li L;
XX
DR   WPI; 1997-289308/26.
DR   N-PSDB; AAT69575.
XX
PT   Nucleic acid encoding tumour susceptibility protein TSG101 - useful
PT   to develop products for cancer detection, prophylaxis, therapy and
PT   predisposition determination
XX
PS   Example 1; Page 73-74; 89pp; English.
XX
CC   This polypeptide sequence comprises a human tumour susceptibility
CC   protein, designated TSG101, that can function as a tumour
CC   suppressor in oncogenesis.  Its amino acid sequence was deduced
CC   from a cDNA clone (AAT69575) isolated from a human placenta cDNA
CC   library.  It shows 94% identity to human TSG101 (see AAW19110).
CC   Deletions in human TSG101 are associated with the occurrence of
CC   cancers, e.g. breast cancer.  Methods are provided for identifying
CC   homologous or related proteins, for producing compositions that
CC   modulate the expression or function of the TSG101 protein, and for
CC   studying physiological pathways.  Modulation of gene activity in
CC   vivo is used for prophylactic or therapeutic purposes, such as
CC   cancer therapy, identification of cell type based on expression,
CC   etc.
XX
SQ   Sequence   380 AA;

  Query Match           100.0%;  Score 2002;  DB 18;  Length 380;
  Best Local Similarity  100.0%;  Pred. No. 4.3e-146;
  Matches  380;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 60

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120

Qy        121 MIVVFGDEPPVFSRPISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCPY 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 MIVVFGDEPPVFSRPISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCPY 180

Qy        181 PPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRAQ 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 PPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRAQ 240

Qy        241 AELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQSE 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQSE 300

Qy        301 NNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRKQ 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 NNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRKQ 360

Qy        361 FQLRALMQKARKTAGLSDLY 380
              ||||||||||||||||||||
Db        361 FQLRALMQKARKTAGLSDLY 380


RESULT 2
AAW93424
ID   AAW93424 standard; Protein; 390 AA.
XX
AC   AAW93424;
XX
DT   11-JUN-1999  (first entry)
XX
DE   Human NHTS protein.
XX
KW   Human tumour suppressor; NHTS; treatment; diagnosis; lymphoma; cancer;
KW   brain; breast; colon; heart; kidney; ovary; paraganglia; pancreas;
KW   prostate; skin; stomach; thyroid; autoimmune disease; asthma; diabetes;
KW   biliary cirrhosis; Crohn's disease; rheumatoid arthritis.
XX
OS   Homo sapiens.
XX
PN   US5892016-A.
XX
PD   06-APR-1999.
XX
PF   23-JAN-1997;   97US-0786999.
XX
PR   23-JAN-1997;   97US-0786999.
XX
PA   (INCY-) INCYTE PHARM.
XX
PI   Brie SL,  Goli SK;
XX
DR   WPI; 1999-253932/21.
DR   N-PSDB; AAX23168.
XX
PT   Novel human tumor suppressor - useful for the diagnosis or treatment
PT   of lymphoma, cancer, and autoimmune disease
XX
PS   Claim 1; Fig 1A-B; 25pp; English.
XX
CC   This invention describes the isolation of a novel human tumour
CC   suppressor (NHTS). The products of the invention may be used for
CC   the diagnosis or treatment of conditions and diseases which are
CC   associated with expression of NHTS e.g. lymphoma and cancers of the
CC   brain, breast, colon, heart, kidney, ovary, paraganglia, pancreas,
CC   prostate, skin, stomach and thyroid and autoimmune disease e.g. asthma,
CC   biliary cirrhosis, Crohn's disease, diabetes, and rheumatoid arthritis.
```

```
XX
SQ   Sequence   390 AA;

  Query Match             100.0%;  Score 2002;  DB 20;  Length 390;
  Best Local Similarity   100.0%;  Pred. No. 4.4e-146;
  Matches  380;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          11 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 70

Qy          61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          71 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 130

Qy         121 MIVVFGDEPPVFSRPISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCPY 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         131 MIVVFGDEPPVFSRPISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCPY 190

Qy         181 PPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRAQ 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         191 PPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRAQ 250

Qy         241 AELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQSE 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         251 AELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQSE 310

Qy         301 NNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRKQ 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         311 NNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRKQ 370

Qy         361 FQLRALMQKARKTAGLSDLY 380
              ||||||||||||||||||||
Db         371 FQLRALMQKARKTAGLSDLY 390


RESULT 3
AAW19110
ID   AAW19110 standard; Protein; 381 AA.
XX
AC   AAW19110;
XX
DT   10-DEC-1997  (first entry)
XX
DE   Mouse tumour susceptibility protein TSG101.
XX
KW   Tumour susceptibility protein; TSG101; tumorigenesis;
KW   breast cancer; diagnosis; therapy; mouse.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   Region          37..46
FT                   /note= "residues 37-46 resemble a helix-turn-helix
FT                    signature domain"
FT   Region          73..83
FT                   /note= "residues 73-83 resemble a fungal Zn-cys
FT                    bi-nuclear cluster signature "
FT   Modified-site   11
FT                   /note= "potential protein kinase C phosphorylation
FT                    site"
FT   Modified-site   38
FT                   /note= "potential protein kinase C phosphorylation
FT                    site"
FT   Modified-site   85
FT                   /note= "potential protein kinase C phosphorylation
FT                    site"
FT   Modified-site   88
FT                   /note= "potential protein kinase C phosphorylation
FT                    site"
FT   Modified-site   215
FT                   /note= "potential protein kinase C phosphorylation
FT                    site"
FT   Modified-site   225
FT                   /note= "potential protein kinase C phosphorylation
FT                    site"
FT   Modified-site   357
FT                   /note= "potential protein kinase C phosphorylation
FT                    site"
FT   Modified-site   38
FT                   /note= "potential casein kinase II phosphorylation
FT                    site"
FT   Modified-site   210
FT                   /note= "potential casein kinase II phosphorylation
FT                    site"
FT   Modified-site   249
FT                   /note= "potential casein kinase II phosphorylation
FT                    site"
FT   Modified-site   265
FT                   /note= "potential casein kinase II phosphorylation
FT                    site"
FT   Modified-site   290
FT                   /note= "potential casein kinase II phosphorylation
FT                    site"
FT   Modified-site   55
FT                   /note= "potential N-myristoylation site"
FT   Modified-site   156
FT                   /note= "potential N-myristoylation site"
FT   Modified-site   44
FT                   /note= "potential N-glycosylation site"
FT   Modified-site   150
FT                   /note= "potential N-glycosylation site"
FT   Modified-site   297
FT                   /note= "potential N-glycosylation site"
XX
PN   WO9718333-A1.
XX
PD   22-MAY-1997.
XX
PF   15-NOV-1996;   96WO-US18828.
XX
PR   13-JUN-1996;   96US-0670274.
PR   16-NOV-1995;   95US-0006856.
PR   16-JAN-1996;   96US-0585758.
XX
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
XX
PI   Cohen SN,  Li L;
XX
DR   WPI; 1997-289308/26.
DR   N-PSDB; AAT69574.
XX
PT   Nucleic acid encoding tumour susceptibility protein TSG101 - useful
PT   to develop products for cancer detection, prophylaxis, therapy and
PT   predisposition determination
XX
PS   Example 1; Page 67-70; 89pp; English.
XX
CC   This polypeptide sequence comprises a mouse tumour susceptibility
CC   protein, designated TSG101, that can function as a tumour
CC   suppressor in oncogenesis.  Its amino acid sequence was deduced
CC   from a cDNA clone (AAT69574) isolated using a novel random homozygous
CC   knockout approach.  It shows 94% identity to human TSG101 (see
CC   AAW19111).  Deletions in human TSG101 are associated with the
CC   occurrence of cancers, e.g. breast cancer.  Methods are provided
CC   for identifying homologous or related proteins, for producing
CC   compositions that modulate the expression or function of the TSG101
CC   protein, and for studying physiological pathways.  Modulation of
CC   gene activity in vivo is used for prophylactic or therapeutic
CC   purposes, such as cancer therapy, identification of cell type based
CC   on expression, etc.
XX
SQ   Sequence   381 AA;

  Query Match             94.9%;  Score 1900.5;  DB 18;  Length 381;
  Best Local Similarity   94.5%;  Pred. No. 2.9e-138;
```

```
  Matches  360;  Conservative   11;  Mismatches    9;  Indels    1;  Gaps    1;

Qy          1 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 60
              |:||||||||||:|||| :||||||||||||||||||:|||||||| |||| |||
Db          1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNI 60

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120
              ||||||||||||||||||||||||||||||||||||||||||:||||:|:|| |||:
Db         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQI 120

Qy        121 MIVVFGDEPPVFSRP-ISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCP 179
              |||:||:|||||||| :|||||| |||||||||||||| ||| ||||||||||||||||
Db        121 MIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCP 180

Qy        180 YPPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRA 239
              ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        181 YPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGA 240

Qy        240 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 300

Qy        300 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 360

Qy        360 QFQLRALMQKARKTAGLSDLY 380
              |||||||||||||||||||||
Db        361 QFQLRALMQKARKTAGLSDLY 381


RESULT 4
AAW93425
ID   AAW93425 standard; Protein; 381 AA.
XX
AC   AAW93425;
XX
DT   11-JUN-1999  (first entry)
XX
DE   Mouse tsg101 protein.
XX
KW   Human tumour suppressor; NHTS; treatment; diagnosis; lymphoma; cancer;
KW   brain; breast; colon; heart; kidney; ovary; paraganglia; pancreas;
KW   prostate; skin; stomach; thyroid; autoimmune disease; asthma; diabetes;
KW   biliary cirrhosis; Crohn's disease; rheumatoid arthritis; mouse; tsg101.
XX
OS   Mus sp.
XX
PN   US5892016-A.
XX
PD   06-APR-1999.
XX
PF   23-JAN-1997;   97US-0786999.
XX
PR   23-JAN-1997;   97US-0786999.
XX
PA   (INCY-) INCYTE PHARM.
XX
PI   Brie SL,  Goli SK;
XX
DR   WPI; 1999-253932/21.
XX
PT   Novel human tumor suppressor - useful for the diagnosis or treatment
PT   of lymphoma, cancer, and autoimmune disease
XX
PS   Disclosure; Fig 2; 25pp; English.
XX
CC   This invention describes the isolation of a novel human tumour
CC   suppressor (NHTS). The products of the invention may be used for
CC   the diagnosis or treatment of conditions and diseases which are
CC   associated with expression of NHTS e.g. lymphoma and cancers of the
CC   brain, breast, colon, heart, kidney, ovary, paraganglia, pancreas,
CC   prostate, skin, stomach and thyroid and autoimmune disease e.g. asthma,
CC   biliary cirrhosis, Crohn's disease, diabetes, and rheumatoid arthritis.
XX
SQ   Sequence   381 AA;

  Query Match             94.9%;  Score 1900.5;  DB 20;  Length 381;
  Best Local Similarity   94.5%;  Pred. No. 2.9e-138;
  Matches  360;  Conservative   11;  Mismatches    9;  Indels    1;  Gaps    1;

Qy          1 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 60
              |:||||||||||:|||| :||||||||||||||||||:|||||||| |||| |||
Db          1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNI 60

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120
              ||||||||||||||||||||||||||||||||||||||||||:||||:|:|| |||:
Db         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQI 120

Qy        121 MIVVFGDEPPVFSRP-ISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCP 179
              |||:||:|||||||| :|||||| |||||||||||||| ||| ||||||||||||||||
Db        121 MIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCP 180

Qy        180 YPPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRA 239
              ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        181 YPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGA 240

Qy        240 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 300

Qy        300 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 360

Qy        360 QFQLRALMQKARKTAGLSDLY 380
              |||||||||||||||||||||
Db        361 QFQLRALMQKARKTAGLSDLY 381


RESULT 5
ABP41729
ID   ABP41729 standard; Protein; 307 AA.
XX
AC   ABP41729;
XX
DT   22-AUG-2002  (first entry)
XX
DE   Human ovarian antigen HSPSE88, SEQ ID NO:2861.
XX
KW   Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
KW   ovarian cancer; breast cancer; tumour; reproductive system disorder;
KW   infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
KW   PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;
KW   inflammatory condition; immune disorder; blood disorder;
KW   cardiovascular disorder; respiratory disorder; neurological disorder;
KW   gastrointestinal disorder; urinary system disorder; drug screening;
KW   gene therapy; chromosome mapping; forensic analysis;
KW   antibody preparation; cytostatic; immunomodulatory; neuroprotective;
KW   antiinflammatory; gynaecological; reproductive; chromosome 11p15.
XX
OS   Homo sapiens.
XX
PN   WO200200677-A1.
XX
PD   03-JAN-2002.
XX
PF   07-JUN-2001; 2001WO-US18569.
XX
PR   07-JUN-2000; 2000US-209467P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Birse CE,  Rosen CA;
XX
DR   WPI; 2002-147878/19.
```

DR N-PSDB; ABQ54806.
XX
PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,
PT useful in the prevention, treatment and diagnosis of cancer (e.g.
PT ovarian cancer), immune disorders, cardiovascular disorders and
PT neurological diseases -
XX
PS Claim 11; SEQ ID No 2861; 2922pp; English.
XX
CC The invention relates to 2175 novel human ovarian antigens (ABP41054-
CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
CC encompasses polypeptides 90% identical and polynucleotides 95% identical
CC to the sequences of the invention. The invention additionally relates to
CC recombinant vectors and host cells comprising human ovarian antigen
CC polynucleotides, antibodies against human ovarian antigens, and the use
CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
CC treating, prognosing or preventing various ovary and/or breast-related
CC disorders. Such conditions include ovarian cancer and breast cancer, and
CC metastatic tumours of ovarian or breast origin, reproductive system
CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
CC vaginitis), immune disorders (e.g., congenital and acquired
CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
CC respiratory disorders, neurological disorders, gastrointestinal disorders
CC and urinary system disorders. Ovarian antigen polypeptides and
CC polynucleotides may also be used in screening for compounds which
CC modulate ovarian antigen expression or activity. The polynucleotides may
CC further be used for gene therapy, chromosome mapping, in the
CC identification of individuals and in forensic analysis, and the
CC polypeptides may be used as food additives or to prepare antibodies
CC useful in disease diagnosis, drug targeting and phenotyping. The present
CC sequence represents a human ovarian antigen of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 307 AA;

Query Match 70.1%; Score 1404; DB 23; Length 307;
Best Local Similarity 100.0%; Pred. No. 4.1e-100;
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        110 PQSDLLGLIQVMIVVFGDEPPVFSRPISASYPPYQATGPPNTSYMPGMPGGISPYPSGYP 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         37 PQSDLLGLIQVMIVVFGDEPPVFSRPISASYPPYQATGPPNTSYMPGMPGGISPYPSGYP 96

Qy        170 PNPSGYPGCPYPPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLR 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         97 PNPSGYPGCPYPPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLR 156

Qy        230 WRMKEEMDRAQAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELS 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        157 WRMKEEMDRAQAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELS 216

Qy        290 SALEKMENQSENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVF 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        217 SALEKMENQSENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVF 276

Qy        350 LKHVRLLSRKQFQLRALMQKARKTAGLSDLY 380
              ||||||||||||||||||||||||||||||
Db        277 LKHVRLLSRKQFQLRALMQKARKTAGLSDLY 307
```

RESULT 6
AAE09328
ID AAE09328 standard; Protein; 237 AA.
XX
AC AAE09328;
XX
DT 19-NOV-2001 (first entry)
XX
DE Human intracellular regulatory molecule, tsg101.
XX
KW Human; intracellular regulator; cell division; proliferation; therapy;
KW cancer; infection; wound; developmental abnormality; metabolic problem;
KW cytostatic; antibacterial; vulnerary; transcription factor; tsg101.
XX
OS Homo sapiens.
XX
PN US6274312-B1.
XX
PD 14-AUG-2001.
XX
PF 10-DEC-1997; 97US-0999774.
XX
PR 11-DEC-1996; 96US-0032818.
XX
PA (SCHE) SCHERING CORP.
XX
PI Gish KC, Seghezzi W, Shanahan F, Lees EM, McClanahan TK;
XX
DR WPI; 2001-535086/59.
DR N-PSDB; AAD16104.
XX
PT New genes encoding intracellular regulatory molecules, useful for
PT regulating cell division and proliferation (e.g. tumor cells),
PT particularly for treating cancer, infections, wounds, or developmental
PT or metabolic abnormalities -
XX
PS Example 2; Column 29-32; 62pp; English.
XX
CC The present sequence is a human intracellular regulatory
CC molecule, tsg101. The tsg101 is a transcription factor. The
CC polynucleotides encoding intracellular regulatory molecules are
CC useful for regulating cell division and proliferation of various
CC cell types, including tumour cells. Specifically, they are also
CC useful for treating cancer, infections, wounds, developmental
CC abnormalities or metabolic problems.
XX
SQ Sequence 237 AA;

Query Match 55.8%; Score 1117; DB 22; Length 237;
Best Local Similarity 98.0%; Pred. No. 3.8e-78;
Matches 201; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```
Qy          1 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         11 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 70

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         71 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 130

Qy        121 MIVVFGDEPPVFSRPISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCPY 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        131 MIVVFGDEPPVFSRPISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCPY 190

Qy        181 PPGGPYPATTSSQYPSQPPVTTVGP 205
              |||||||||||||||||||  : |
Db        191 PPGGPYPATTSSQYPSQPPVPLLVP 215
```

RESULT 7
ABB64607
ID ABB64607 standard; Protein; 331 AA.
XX
AC ABB64607;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 20613.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;

```
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL08710.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 20613; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   331 AA;

  Query Match             40.4%;  Score 809.5;  DB 22;  Length 331;
  Best Local Similarity   50.0%;  Pred. No. 3.1e-54;
  Matches  168;  Conservative   51;  Mismatches   86;  Indels   31;  Gaps    7;

Qy         67 LDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQVMIVVFG 126
Db          1 MDTHPQNAPMCFVKPTPTMQIKVSMYVDHNGKVYLPYLHDWQPHSSDLLSLIQVMIVTFG 60

Qy        127 DEPPVFSRPISASYPPYQATGPPNTSYM--PGMPGGIS---PYP-------SGYPPNPSG 174
Db         61 DHPPVYSKPKEQIAAPY-----PTNSYMPQPGAPGGSNSFLPYPTAGGAGGSNFPPYPTG 115

Qy        175 YPGCPYPP-------GGPYPA------TTSSQYPSQPPVTTVGPSRDGTISEDTIRASLI 221
Db        116 SNVGPYPPTPAGPAGGSGYPAYPNFIQPTAGGYPPAAGYNPSNPSSTGTITEEHIKASII 175

Qy        222 SAVSDKLRWRMKEEMDRAQAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELL 281
Db        176 SAIDDKLRRRVQEKVNQYQAEIETLNRTKQELLEGSAKIDAIIERLEREHIDMQKNISIL 235

Qy        282 KKKDEELSSALEKMENQSENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRR 341
Db        236 KDKEQELEKALEDLESAEAINP-DEAVTTTAPLYRQLLNAYADEAATEDAIYYLGEGLRG 294

Qy        342 GVIDLDVFLKHVRLLSRKQFQLRALMQKARKTAGLS 377
Db        295 GVIDLETFLKHVRQLSRKQFILRATMQKCRQKAGLA 330


RESULT 8
AAG06370
ID   AAG06370 standard; Protein; 398 AA.
XX
```

```
AC   AAG06370;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 3119.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
```

PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 22.7%; Score 455; DB 21; Length 398;
Best Local Similarity 32.1%; Pred. No. 8.4e-27;
Matches 117; Conservative 69; Mismatches 146; Indels 32; Gaps 10;

```
Qy         16 VNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNIPICLWLLDTYPYNPP 75
              :|:|: |  |:|   |:: ||| |  |:   ||||:|: | ||||||: :|||::|| :||
Db         42 LNLISSYPSLEPKTASFMHNDGRSVNLLQADGTIPMPFHGVTYNIPVIIWLLESYPRHPP 101

Qy         76 ICFVKPTSSMTIK-TGKHVDANGKIYLPYLHEWKHPQSDLLGLIQVMIVVFGDEPPVFSR 134
                :| ||: | ||    ||  :| : ||||  | :| |:|: |:  :   |  :||::||
Db        102 CVYVNPTADMIIKRPHAHVTPSGLVSLPYLQNWVYPSSNLVDLVSDLSAAFARDPPLYSR 161

Qy        135 PISASYPPYQATGPPNT--SYMPGMPGGISPYPSGYPPNPSGYPGCPYPPGGPYPATTSS 192
                      ||     ||    | :   |      |  :||:| |        ||
Db        162 RRPQPPPP----SPPTVYDSSLSRPPSADQSLPRPFPPSPYG--------GGVSRVQVQH 209

Qy        193 QYPSQPPVTTVGPSRDGTISE--DTIRASLISAVSDKLRWRMKEEMDRAQAELNALKRTE 250
               :  |       :  |:: : : : |:|   : |   ||:  ||   ||| |
Db        210 VHHQQQSDDAAEVFKRNAINKMVEMVHSDLVS--MRRAREAEAEELLSLQA---GLKRRE 264
```

```
Qy        251 EDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQSEN---NDIDEV 307
              ::|  |   |:|||    :|  ::: ::::    : | | : : : :::|    |:|
Db        265 DELNIG---LKEMV----EEKETLEQQLQIISMNTDILDSWVRENQGKTKNLVDLDVDNA 317

Qy        308 IIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRKQFQLRALM 367
                  | ||:|   | : |||| |: | :: : ||:  | :|::||||||:||  ||
Db        318 FECGDTLSKQMLECTALDLAIEDAIYSLDKSFQDGVVPFDQYLRNVRLLSREQFFHRATG 377

Qy        368 QKAR 371
               | |
Db        378 SKVR 381


RESULT 9
AAG06369
ID   AAG06369 standard; Protein; 412 AA.
XX
AC   AAG06369;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 3118.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145218.
PR   23-JUL-1999;   99US-0145224.
PR   26-JUL-1999;   99US-0145276.
PR   27-JUL-1999;   99US-0145913.
PR   27-JUL-1999;   99US-0145918.
PR   27-JUL-1999;   99US-0145919.
PR   28-JUL-1999;   99US-0145951.
PR   02-AUG-1999;   99US-0146386.
PR   02-AUG-1999;   99US-0146388.
PR   02-AUG-1999;   99US-0146389.
PR   03-AUG-1999;   99US-0147038.
PR   04-AUG-1999;   99US-0147204.
PR   04-AUG-1999;   99US-0147302.
PR   05-AUG-1999;   99US-0147192.
PR   05-AUG-1999;   99US-0147260.
PR   06-AUG-1999;   99US-0147303.
```

```
PR   06-AUG-1999;   99US-0147416.
PR   09-AUG-1999;   99US-0147493.
PR   09-AUG-1999;   99US-0147935.
PR   10-AUG-1999;   99US-0148171.
PR   11-AUG-1999;   99US-0148319.
PR   12-AUG-1999;   99US-0148341.
PR   13-AUG-1999;   99US-0148565.
PR   13-AUG-1999;   99US-0148684.
PR   16-AUG-1999;   99US-0149368.
PR   17-AUG-1999;   99US-0149175.
PR   18-AUG-1999;   99US-0149426.
PR   20-AUG-1999;   99US-0149722.
PR   20-AUG-1999;   99US-0149723.
PR   20-AUG-1999;   99US-0149929.
PR   23-AUG-1999;   99US-0149902.
PR   23-AUG-1999;   99US-0149930.
PR   25-AUG-1999;   99US-0150566.
PR   26-AUG-1999;   99US-0150884.
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match          22.7%;  Score 455;  DB 21;  Length 412;
  Best Local Similarity  32.1%;  Pred. No. 8.8e-27;
  Matches  117;  Conservative   69;  Mismatches  146;  Indels   32;  Gaps   10;

Qy        16 VNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNIPICLWLLDTYPYNPP 75
             :|:|: |  |:|   |:: ||| |  |:   ||||:|: | |||||: :|||::|| :||
Db        56 LNLISSYPSLEPKTASFMHNDGRSVNLLQADGTIPMPFHGVTYNIPVIIWLLESYPRHPP 115

Qy        76 ICFVKPTSSMTIK-TGKHVDANGKIYLPYLHEWKHPQSDLLGLIQVMIVVFGDEPPVFSR 134
               :| ||: | ||    ||  :| : ||||  | :| |:|: |:  :   |  :||::||
Db       116 CVYVNPTADMIIKRPHAHVTPSGLVSLPYLQNWVYPSSNLVDLVSDLSAAFARDPPLYSR 175

Qy       135 PISASYPPYQATGPPNT--SYMPGMPGGISPYPSGYPPNPSGYPGCPYPPGGPYPATTSS 192
                 ||    ||   | :   |      |  :||:| |       ||
Db       176 RRPQPPPP----SPPTVYDSSLSRPPSADQSLPRPFPPSPYG--------GGVSRVQVQH 223

Qy       193 QYPSQPPVTTVGPSRDGTISE--DTIRASLISAVSDKLRWRMKEEMDRAQAELNALKRTE 250
             : |         :   |::  : : : |:|    : |    ||:   ||    ||| |
Db       224 VHHQQQSDDAAEVFKRNAINKMVEMVHSDLVS--MRRAREAEAEELLSLQA---GLKRRE 278

Qy       251 EDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQSEN---NDIDEV 307
             ::|  |   |:|||    :|   ::: :::::    : | | : : : :::|    |:|
Db       279 DELNIG---LKEMV----EEKETLEQQLQIISMNTDILDSWVRENQGKTKNLVDLDVDNA 331

Qy       308 IIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRKQFQLRALM 367
                 | ||:|   | : |||| |: | :: : ||:  | :|::|||||||:|| ||
Db       332 FECGDTLSKQMLECTALDLAIEDAIYSLDKSFQDGVVPFDQYLRNVRLLSREQFFHRATG 391

Qy       368 QKAR 371
              | |
Db       392 SKVR 395


RESULT 10
AAG01689
ID   AAG01689 standard; Protein; 90 AA.
XX
AC   AAG01689;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein, SEQ ID NO: 5770.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   N-PSDB; AAC01695.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 13; SEQ ID 5770; 71pp + CD-ROM; English.
XX
CC   The present sequence is a polypeptide encoded by one of a large number
CC   of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC   were prepared from total human RNAs or polyA+ RNAs derived from 30
CC   different tissues. EST sequences usually correspond mainly to the 3'
CC   untranslated region (UTR) of the mRNA because they are often obtained
CC   from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC   isolating cDNA sequences derived from the 5' ends of mRNAs and even in
```

```
CC   those cases where longer cDNA sequences have been obtained, the full 5'
CC   UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC   ends and can therefore be used to obtain full length cDNAs and genomic
CC   DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC   chromosome mapping procedures. They are used to obtain upstream
CC   regulatory sequences and to design expression and secretion vectors.
XX
SQ   Sequence   90 AA;

  Query Match             21.7%;  Score 434;  DB 21;  Length 90;
  Best Local Similarity   100.0%;  Pred. No. 4.7e-26;
  Matches   80;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         11 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 70

Qy         61 PICLWLLDTYPYNPPICFVK 80
              ||||||||||||||||||||
Db         71 PICLWLLDTYPYNPPICFVK 90


RESULT 11
AAG06371
ID   AAG06371 standard; Protein; 340 AA.
XX
AC   AAG06371;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 3120.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145218.
PR   23-JUL-1999;   99US-0145224.
PR   26-JUL-1999;   99US-0145276.
PR   27-JUL-1999;   99US-0145913.
PR   27-JUL-1999;   99US-0145918.
PR   27-JUL-1999;   99US-0145919.
PR   28-JUL-1999;   99US-0145951.
```

PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 21.2%; Score 424; DB 21; Length 340;
Best Local Similarity , 32.2%; Pred. No. 1.7e-24;
Matches 111; Conservative 63; Mismatches 139; Indels 32; Gaps 10;

```
Qy         35 NDGSSRELMNLTGTIPVPYRGNTYNIPICLWLLDTYPYNPPICFVKPTSSMTIK-TGKHV 93
              ||| |  |:   ||||:|: | ||||||: :|||::|| :||  :| ||: | ||    ||
Db          3 NDGRSVNLLQADGTIPMPFHGVTYNIPVIIWLLESYPRHPPCVYVNPTADMIIKRPHAHV 62

Qy         94 DANGKIYLPYLHEWKHPQSDLLGLIQVMIVVFGDEPPVFSRPISASYPPYQATGPPNT-- 151
                :| : |||| | :| |:|: |:  :   |  :||::||      ||     ||
Db         63 TPSGLVSLPYLQNWVYPSSNLVDLVSDLSAAFARDPPLYSRRRPQPPPP----SPPTVYD 118

Qy        152 SYMPGMPGGISPYPSGYPPNPSGYPGCPYPPGGPYPATTSSQYPSQPPVTTVGPSRDGTI 211
              | :   |      |  :||:| |        ||         :  |        :   |
Db        119 SSLSRPPSADQSLPRPFPPSPYG--------GGVSRVQVQHVHHQQQSDDAAEVFKRNAI 170

Qy        212 SE--DTIRASLISAVSDKLRWRMKEEMDRAQAELNALKRTEEDLKKGHQKLEEMVTRLDQ 269
              ::  : : : |:|    : |    ||:   ||    ||| |::|  |   |:|||    :
Db        171 NKMVEMVHSDLVS--MRRAREAEAEELLSLQA---GLKRREDELNIG---LKEMV----E 218

Qy        270 EVAEVDKNIELLKKKDEELSSALEKMENQSEN---NDIDEVIIPTAPLYKQILNLYAEEN 326
              |   ::: ::::    : | | : : :::|    |:|        | ||:|   | :
Db        219 EKETLEQQLQIISMNTDILDSWVRENQGKTKNLVDLDVDNAFECGDTLSKQMLECTALDL 278

Qy        327 AIEDTIFYLGEALRRGVIDLLDVFLKHVRLLSRKQFQLRALMQKAR 371
              |||| |: | :: : ||:  | :|::|||||:||  ||   | |
Db        279 AIEDAIYSLDKSFQDGVVPFDQYLRNVRLLSREQFFHRATGSKVR 323
```

RESULT 12
AAG13898
ID AAG13898 standard; Protein; 322 AA.
XX
AC AAG13898;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 13559.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.

PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 19.0%; Score 380; DB 21; Length 322;
Best Local Similarity 31.0%; Pred. No. 3.8e-21;
Matches 102; Conservative 62; Mismatches 133; Indels 32; Gaps 10;

```
Qy         51 VPYRGNTYNIPICLWLLDTYPYNPPICFVKPTSSMTIK-TGKHVDANGKIYLPYLHEWKH 109
              :|: | |||||: :|||::|| :||  :| ||: | ||    ||  :| : ||||  | :
Db          1 MPFHGVTYNIPVIIWLLESYPRHPPCVYVNPTADMIIKRPHAHVTPSGLVSLPYLQNWVY 60

Qy        110 PQSDLLGLIQVMIVVFGDEPPVFSRPISASYPPYQATGPPNT--SYMPGMPGGISPYPSG 167
              | |:|: |:  :   |  :||::||      ||     ||    | :   |      |
Db         61 PSSNLVDLVSDLSAAFARDPPLYSRRRPQPPPP----SPPTVYDSSLSRPPSADQSLPRP 116

Qy        168 YPPNPSGYPGCPYPPGGPYPATTSSQYPSQPPVTTVGPSRDGTISE--DTIRASLISAVS 225
              :||:| |         ||         :  |         :   |::  : : : |:|
Db        117 FPPSPYG--------GGVSRVQVQHVHHQQQSDDAAEVFKRNAINKMVEMVHSDLVS--M 166

Qy        226 DKLRWRMKEEMDRAQAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKD 285
               : |    ||:   ||    ||| |::|  |   |:|||    :|   ::: ::::
Db        167 RRAREAEAEELLSLQA---GLKRREDELNIG---LKEMV----EEKETLEQQLQIISMNT 216

Qy        286 EELSSALEKMENQSEN---NDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRG 342
              : | | : : :::|    |:|        | ||:|   | : |||| |: | :: : |
Db        217 DILDSWVRENQGKTKNLVDLDVDNAFECGDTLSKQMLECTALDLAIEDAIYSLDKSFQDG 276

Qy        343 VIDLDVFLKHVRLLSRKQFQLRALMQKAR 371
              |:  | :|::|||||:||  ||   | |
Db        277 VVPFDQYLRNVRLLSREQFFHRATGSKVR 305
```

RESULT 13
AAB93473
ID AAB93473 standard; Protein; 341 AA.
XX
AC AAB93473;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:12750.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 12750; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 341 AA;

Query Match 18.5%; Score 371; DB 22; Length 341;
Best Local Similarity 60.0%; Pred. No. 2e-20;
Matches 63; Conservative 21; Mismatches 21; Indels 0; Gaps 0;

```
Qy         29 LDSYVFNDGSSRELMNLTGTIPVPYRGNTYNIPICLWLLDTYPYNPPICFVKPTSSMTIK 88
              :|:||| | | ::|:| |||||| |:||||||||  |:||::|: ||||:|||::| |
Db          1 MDTYVFKDSSQKDLLNFTGTIPVMYQGNTYNIPIRFWILDSHPFAPPICFLKPTANMGIL 60

Qy         89 TGKHVDANGKIYLPYLHEWKHPQSDLLGLIQVMIVVFGDEPPVFS 133
               |||||| |:|||||  | ||:| ::|||: ||  | :| |::|
Db         61 VGKHVDAQGRIYLPYLQNWSHPKSVIVGLIKEMIAKFQEELPMYS 105
```

RESULT 14
AAB53717
ID AAB53717 standard; Protein; 146 AA.
XX
AC AAB53717;
XX
DT 09-MAR-2001 (first entry)
XX
DE Human colon cancer antigen protein sequence SEQ ID NO:1257.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW identification; cytostatic; cardioactive; neuroprotective; vulnerary;
KW immunomodulatory; muscular; gynaecological; gastrointestinal;
KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;
KW neural disorder; immune system disorder; muscular disorder;
KW reproductive disorder; gastrointestinal disorder; renal disorder;
KW infectious disease; cardiovascular disorder.
XX
OS Homo sapiens.
XX

PN WO200055351-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05883.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2000-587534/55.
DR N-PSDB; AAC98474.
XX
PT Colon cancer associated gene sequences, referred to as colon cancer
PT antigens, useful for the treatment, prevention, and diagnosis of colon
PT disorders such as colon cancer -
XX
PS Claim 11; Page 1845; 2104pp; English.
XX
CC AAC97991 to AAC98763 encode the human colon cancer associated proteins,
CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
CC vulnerary, nephrotropic, antiinfective and antibacterial activities, and
CC can be used in gene therapy. The colon cancer antigen polynucleotides,
CC proteins and antibodies to the proteins are useful for the prevention,
CC treatment and diagnosis of colon disorders, such as colon cancer. The
CC polynucleotides may be used in diagnostics and research, such as for
CC chromosome identification, and as hybridisation probes. The proteins
CC may also be used to prevent diseases such as neural disorders, immune
CC system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, wounds, renal disorders, infectious
CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
CC AAB54007 represent sequences used in the exemplification of the present
CC invention.
XX
SQ Sequence 146 AA;

Query Match 17.3%; Score 346; DB 21; Length 146;
Best Local Similarity 95.5%; Pred. No. 5.5e-19;
Matches 63; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy         33 VFNDGSSRELMNLTGTIPVPYRGNTYNIPICLWLLDTYPYNPPICFVKPTSSMTIKTGKH 92
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         69 VFNDGSSRELMNLTGTIPVPYRGNTYNIPICLWLLDTYPYNPPICFVKPTSSMTIKTGKH 128

Qy         93 VDANGK 98
              ||   |
Db        129 VDXPKK 134
```

RESULT 15
AAU23116
ID AAU23116 standard; Protein; 452 AA.
XX
AC AAU23116;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human enzyme polypeptide #202.
XX
KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;
KW nephrotropic; anticoagulant.
XX
OS Homo sapiens.
XX
PN WO200155301-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01239.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.

PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465566/50.
DR N-PSDB; AAS40986.
XX
PT Novel polypeptides and polynucleotides useful for diagnosing,
PT preventing, treating neural, immune system, muscular, reproductive,
PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous
PT diseases -
XX
PS Claim 11; SEQ ID No 1112; 1180pp; English.
XX
CC The present invention relates to the isolation of novel human enzyme
CC polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences
CC encoding them. The enzyme polypeptides of the invention may comprise the
CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
CC isomerases or ligases. The sequences of the invention are useful in the
CC diagnosis, treatment, prevention and/or prognosis of a wide range of
CC disorders including hyperproliferative disorders (e.g. cancer),
CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders
CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders
CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
CC blood-related disorders (e.g. haemophilia), reproductive disorders
CC (e.g. infertility) and infectious disorders (e.g. Influenza). The
CC polynucleotides of the invention can also be used in gene therapy.
CC AAU22915-AAU23814 represent the novel human enzyme polypeptides of the
CC invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 452 AA;

Query Match 16.5%; Score 331; DB 22; Length 452;
Best Local Similarity 46.6%; Pred. No. 3.6e-17;
Matches 62; Conservative 20; Mismatches 29; Indels 22; Gaps 1;

```
Qy          1 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 60
              :  |||:||||| |  ||   :   |  :|:|                      |||||||
Db         14 LXGKYKFRDLTVEELRNVNVFFPHFKYSMDTY--------------------GNTYNI 51

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120
              ||  |:||::|: |||||:|||::| |  |||||| |:||||||  | ||:| ::|||:
Db         52 PIRFWILDSHPFAPPICFLKPTANMGILVGKHVDAQGRIYLPYLQNWSHPKSVIVGLIKE 111

Qy        121 MIVVFGDEPPVFS 133
              ||  | :| |::|
Db        112 MIAKFQEELPMYS 124
```

Search completed: June 3, 2003, 14:07:50
Job time : 52.4323 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: June 3, 2003, 14:06:00 ; Search time 12.9829 Seconds
(without alignments)
861.186 Million cell updates/sec

Title: US-09-804-690-4
Perfect score: 2002
Sequence: 1 MVSKYKYRDLTVRETVNVIT.....FQLRALMQKARKTAGLSDLY 380

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2002	100.0	380	1	US-08-585-758A-4	Sequence 4, Appli
2	2002	100.0	380	1	US-08-977-818-4	Sequence 4, Appli
3	2002	100.0	380	2	US-08-670-274B-4	Sequence 4, Appli
4	2002	100.0	380	4	US-09-146-187-4	Sequence 4, Appli
5	2002	100.0	390	2	US-08-786-999-1	Sequence 1, Appli
6	1900.5	94.9	381	1	US-08-585-758A-2	Sequence 2, Appli
7	1900.5	94.9	381	1	US-08-977-818-2	Sequence 2, Appli
8	1900.5	94.9	381	2	US-08-670-274B-2	Sequence 2, Appli
9	1900.5	94.9	381	2	US-08-786-999-3	Sequence 3, Appli
10	1900.5	94.9	381	4	US-09-146-187-2	Sequence 2, Appli
11	1117	55.8	237	4	US-08-999-774A-8	Sequence 8, Appli
12	160	8.0	466	3	US-08-526-136-13	Sequence 13, Appl
13	141	7.0	905	2	US-08-574-959A-9	Sequence 9, Appli
14	141	7.0	905	4	US-09-357-014-9	Sequence 9, Appli
15	141	7.0	1135	2	US-08-574-959A-7	Sequence 7, Appli
16	141	7.0	1135	4	US-09-357-014-7	Sequence 7, Appli
17	130	6.5	543	4	US-09-535-008-63	Sequence 63, Appl
18	130	6.5	577	4	US-09-535-008-61	Sequence 61, Appl
19	130	6.5	1646	4	US-09-535-008-67	Sequence 67, Appl
20	130	6.5	1647	4	US-09-535-008-2	Sequence 2, Appli
21	130	6.5	1649	4	US-09-535-008-75	Sequence 75, Appl
22	130	6.5	1650	4	US-09-535-008-71	Sequence 71, Appl
23	130	6.5	1678	4	US-09-535-008-69	Sequence 69, Appl
24	130	6.5	1679	4	US-09-535-008-65	Sequence 65, Appl
25	130	6.5	1681	4	US-09-535-008-77	Sequence 77, Appl
26	130	6.5	1682	4	US-09-535-008-73	Sequence 73, Appl
27	127	6.3	1185	4	US-09-041-886-23	Sequence 23, Appl
28	125.5	6.3	330	1	US-08-642-255-32	Sequence 32, Appl
29	125.5	6.3	408	1	US-07-609-716-65	Sequence 65, Appl
30	125.5	6.3	408	4	US-08-475-411A-65	Sequence 65, Appl
31	125.5	6.3	408	4	US-08-478-029A-65	Sequence 65, Appl
32	124.5	6.2	1162	2	US-08-728-323A-2	Sequence 2, Appli
33	124.5	6.2	1162	4	US-09-298-568-2	Sequence 2, Appli
34	124.5	6.2	1274	4	US-09-095-443-2	Sequence 2, Appli
35	123.5	6.2	380	2	US-09-026-587-4	Sequence 4, Appli
36	123.5	6.2	380	2	US-09-227-420-4	Sequence 4, Appli
37	123	6.1	1248	2	US-09-080-897-2	Sequence 2, Appli
38	123	6.1	1248	4	US-09-323-735-2	Sequence 2, Appli
39	122	6.1	1315	3	US-08-899-595-3	Sequence 3, Appli
40	119.5	6.0	503	3	US-08-526-136-2	Sequence 2, Appli
41	119.5	6.0	505	3	US-08-526-136-4	Sequence 4, Appli
42	118.5	5.9	1255	2	US-09-080-897-4	Sequence 4, Appli
43	118.5	5.9	1255	3	US-08-899-595-1	Sequence 1, Appli
44	118.5	5.9	1255	4	US-09-323-735-4	Sequence 4, Appli
45	118	5.9	214	1	US-08-217-327-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-08-585-758A-4
; Sequence 4, Application US/08585758A
; Patent No. 5679523
; GENERAL INFORMATION:
; APPLICANT: LI, Limin
; APPLICANT: COHEN, Stanley N.
; TITLE OF INVENTION: METHOD FOR CONCURRENT DISRUPTION OF
; TITLE OF INVENTION: EXPRESSION OF MULTIPLE ALLELES OF MAMMALIAN GENES
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: FOUR EMBARCADERO CENTER, SUITE 3400
; CITY: SAN FRANCISCO
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/585,758A
; FILING DATE: 12-JAN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertram I
; REGISTRATION NUMBER: 20015
; REFERENCE/DOCKET NUMBER: A62783/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 380 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-585-758A-4

Query Match 100.0%; Score 2002; DB 1; Length 380;
Best Local Similarity 100.0%; Pred. No. 3.8e-155;
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 60

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120
```

```
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120

Qy         121 MIVVFGDEPPVFSRPISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCPY 180
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 MIVVFGDEPPVFSRPISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCPY 180

Qy         181 PPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRAQ 240
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 PPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRAQ 240

Qy         241 AELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQSE 300
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 AELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQSE 300

Qy         301 NNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRKQ 360
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 NNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRKQ 360

Qy         361 FQLRALMQKARKTAGLSDLY 380
                    ||||||||||||||||||||
Db         361 FQLRALMQKARKTAGLSDLY 380


RESULT 2
US-08-977-818-4
; Sequence 4, Application US/08977818
; Patent No. 5807995
;  GENERAL INFORMATION:
;    APPLICANT:  LI, Limin
;    APPLICANT:  COHEN, Stanley N
;    TITLE OF INVENTION:  MAMMALIAN TUMOR SUSCEPTIBILITY GENES AND
;    TITLE OF INVENTION:  THEIR USES
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FISH AND RICHARDSON, P.C.
;      STREET:  2200 SAND HILL ROAD
;      CITY:  MENLO PARK
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94025
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/977,818
;      FILING DATE:  25-NOV-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/670,274
;      FILING DATE:  June 13, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SHERWOOD, Pamela J.
;      REGISTRATION NUMBER:  36,677
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-781-1989
;      TELEFAX:  415-398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  380 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-977-818-4

  Query Match           100.0%;  Score 2002;  DB 1;  Length 380;
  Best Local Similarity  100.0%;  Pred. No. 3.8e-155;
  Matches  380;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 60
```

```
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 60

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120

Qy        121 MIVVFGDEPPVFSRPISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCPY 180
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 MIVVFGDEPPVFSRPISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCPY 180

Qy        181 PPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRAQ 240
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 PPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRAQ 240

Qy        241 AELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQSE 300
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQSE 300

Qy        301 NNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRKQ 360
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 NNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRKQ 360

Qy        361 FQLRALMQKARKTAGLSDLY 380
                    ||||||||||||||||||||
Db        361 FQLRALMQKARKTAGLSDLY 380


RESULT 3
US-08-670-274B-4
; Sequence 4, Application US/08670274B
; Patent No. 5891668
;  GENERAL INFORMATION:
;    APPLICANT:  LI, Limin
;    APPLICANT:  COHEN, Stanley N
;    TITLE OF INVENTION:  MAMMALIAN TUMOR SUSCEPTIBILITY GENES AND
;    TITLE OF INVENTION:  THEIR USES
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FISH AND RICHARDSON, P.C.
;      STREET:  2200 SAND HILL ROAD
;      CITY:  MENLO PARK
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94025
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/670,274B
;      FILING DATE:  June 13, 1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SHERWOOD, Pamela J.
;      REGISTRATION NUMBER:  36,677
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-781-1989
;      TELEFAX:  415-398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  380 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-670-274B-4

  Query Match           100.0%;  Score 2002;  DB 2;  Length 380;
  Best Local Similarity  100.0%;  Pred. No. 3.8e-155;
  Matches  380;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 60

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120

Qy        121 MIVVFGDEPPVFSRPISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCPY 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 MIVVFGDEPPVFSRPISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCPY 180

Qy        181 PPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRAQ 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 PPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRAQ 240

Qy        241 AELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQSE 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQSE 300

Qy        301 NNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRKQ 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 NNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRKQ 360

Qy        361 FQLRALMQKARKTAGLSDLY 380
              ||||||||||||||||||||
Db        361 FQLRALMQKARKTAGLSDLY 380
```

RESULT 4
US-09-146-187-4
; Sequence 4, Application US/09146187
; Patent No. 6248523
; GENERAL INFORMATION:
; APPLICANT: LI, Limin
; APPLICANT: COHEN, Stanley N
; TITLE OF INVENTION: MAMMALIAN TUMOR SUSCEPTIBILITY GENES AND
; TITLE OF INVENTION: THEIR USES
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH AND RICHARDSON, P.C.
; STREET: 2200 SAND HILL ROAD
; CITY: MENLO PARK
; STATE: CA
; COUNTRY: USA
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/146,187
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/670,274
; FILING DATE: June 13, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: SHERWOOD, Pamela J.
; REGISTRATION NUMBER: 36,677
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 380 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-146-187-4

Query Match 100.0%; Score 2002; DB 4; Length 380;
Best Local Similarity 100.0%; Pred. No. 3.8e-155;
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 60

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120

Qy        121 MIVVFGDEPPVFSRPISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCPY 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 MIVVFGDEPPVFSRPISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCPY 180

Qy        181 PPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRAQ 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 PPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRAQ 240

Qy        241 AELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQSE 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQSE 300

Qy        301 NNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRKQ 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 NNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRKQ 360

Qy        361 FQLRALMQKARKTAGLSDLY 380
              ||||||||||||||||||||
Db        361 FQLRALMQKARKTAGLSDLY 380
```

RESULT 5
US-08-786-999-1
; Sequence 1, Application US/08786999
; Patent No. 5892016
; GENERAL INFORMATION:
; APPLICANT: La Brie, Sam
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN TUMOR
; TITLE OF INVENTION: SUPPRESSOR
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/786,999
; FILING DATE: Filed Herewith
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0199 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 390 amino acids
; TYPE: amino acid

```
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  Colnnot01
;      CLONE:  609476
US-08-786-999-1

  Query Match             100.0%;  Score 2002;  DB 2;  Length 390;
  Best Local Similarity   100.0%;  Pred. No. 3.9e-155;
  Matches  380;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         11 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 70

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         71 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 130

Qy        121 MIVVFGDEPPVFSRPISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCPY 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        131 MIVVFGDEPPVFSRPISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCPY 190

Qy        181 PPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRAQ 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        191 PPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRAQ 250

Qy        241 AELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQSE 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        251 AELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQSE 310

Qy        301 NNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRKQ 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        311 NNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRKQ 370

Qy        361 FQLRALMQKARKTAGLSDLY 380
              ||||||||||||||||||||
Db        371 FQLRALMQKARKTAGLSDLY 390


RESULT 6
US-08-585-758A-2
; Sequence 2, Application US/08585758A
; Patent No. 5679523
;  GENERAL INFORMATION:
;    APPLICANT:  LI, Limin
;    APPLICANT:  COHEN, Stanley N.
;    TITLE OF INVENTION:  METHOD FOR CONCURRENT DISRUPTION OF
;    TITLE OF INVENTION:  EXPRESSION OF MULTIPLE ALLELES OF MAMMALIAN GENES
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  FOUR EMBARCADERO CENTER, SUITE 3400
;      CITY:  SAN FRANCISCO
;      STATE:  CA
;      COUNTRY:  US
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/585,758A
;      FILING DATE:  12-JAN-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rowland, Bertram I
;      REGISTRATION NUMBER:  20015
;      REFERENCE/DOCKET NUMBER:  A62783/BIR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-781-1989
;      TELEFAX:  415-398-3249
```

```
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  381 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-585-758A-2

  Query Match             94.9%;  Score 1900.5;  DB 1;  Length 381;
  Best Local Similarity   94.5%;  Pred. No. 6.9e-147;
  Matches  360;  Conservative   11;  Mismatches    9;  Indels    1;  Gaps    1;

Qy          1 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 60
              |:||||||||||:||||| :|||||||||||||||||:|||||||| |||| |||
Db          1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNI 60

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120
              ||||||||||||||||||||||||||||||||||||||:||||:|:|| |||:
Db         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQI 120

Qy        121 MIVVFGDEPPVFSRP-ISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCP 179
              |||:||:|||||||| :|||||| |||||||||||||| ||| ||||||||||||||||
Db        121 MIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCP 180

Qy        180 YPPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRA 239
              ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        181 YPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGA 240

Qy        240 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 300

Qy        300 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 360

Qy        360 QFQLRALMQKARKTAGLSDLY 380
              ||||||||||||||||||||
Db        361 QFQLRALMQKARKTAGLSDLY 381


RESULT 7
US-08-977-818-2
; Sequence 2, Application US/08977818
; Patent No. 5807995
;  GENERAL INFORMATION:
;    APPLICANT:  LI, Limin
;    APPLICANT:  COHEN, Stanley N
;    TITLE OF INVENTION:  MAMMALIAN TUMOR SUSCEPTIBILITY GENES AND
;    TITLE OF INVENTION:  THEIR USES
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FISH AND RICHARDSON, P.C.
;      STREET:  2200 SAND HILL ROAD
;      CITY:  MENLO PARK
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94025
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/977,818
;      FILING DATE:  25-NOV-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/670,274
;      FILING DATE:  June 13, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SHERWOOD, Pamela J.
;      REGISTRATION NUMBER:  36,677
```

```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-781-1989
;      TELEFAX:  415-398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  381 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-977-818-2

  Query Match             94.9%;  Score 1900.5;  DB 1;  Length 381;
  Best Local Similarity   94.5%;  Pred. No. 6.9e-147;
  Matches  360;  Conservative   11;  Mismatches    9;  Indels    1;  Gaps    1;

Qy          1 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 60
              |:||||||||||||:||||| :|||||||||||||||||||||:|||||||| |||| |||
Db          1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNI 60

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120
              |||||||||||||||||||||||||||||||||||||||||||:||||:|:|| |||:
Db         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQI 120

Qy        121 MIVVFGDEPPVFSRP-ISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCP 179
              |||:||:|||||||| :|||||| ||||||||||||| ||| ||||||||||||||||
Db        121 MIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCP 180

Qy        180 YPPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRA 239
              ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        181 YPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGA 240

Qy        240 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 300

Qy        300 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 360

Qy        360 QFQLRALMQKARKTAGLSDLY 380
              |||||||||||||||||||||
Db        361 QFQLRALMQKARKTAGLSDLY 381


RESULT 8
US-08-670-274B-2
; Sequence 2, Application US/08670274B
; Patent No. 5891668
;  GENERAL INFORMATION:
;    APPLICANT:  LI, Limin
;    APPLICANT:  COHEN, Stanley N
;    TITLE OF INVENTION:  MAMMALIAN TUMOR SUSCEPTIBILITY GENES AND
;    TITLE OF INVENTION:  THEIR USES
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FISH AND RICHARDSON, P.C.
;      STREET:  2200 SAND HILL ROAD
;      CITY:  MENLO PARK
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94025
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/670,274B
;      FILING DATE:  June 13, 1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SHERWOOD, Pamela J.
;      REGISTRATION NUMBER:  36,677
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-781-1989
;      TELEFAX:  415-398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  381 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-670-274B-2

  Query Match             94.9%;  Score 1900.5;  DB 2;  Length 381;
  Best Local Similarity   94.5%;  Pred. No. 6.9e-147;
  Matches  360;  Conservative   11;  Mismatches    9;  Indels    1;  Gaps    1;

Qy          1 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 60
              |:||||||||||||:||||| :|||||||||||||||||||||:|||||||| |||| |||
Db          1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNI 60

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120
              |||||||||||||||||||||||||||||||||||||||||||:||||:|:|| |||:
Db         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQI 120

Qy        121 MIVVFGDEPPVFSRP-ISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCP 179
              |||:||:|||||||| :|||||| ||||||||||||| ||| ||||||||||||||||
Db        121 MIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCP 180

Qy        180 YPPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRA 239
              ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        181 YPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGA 240

Qy        240 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 300

Qy        300 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 360

Qy        360 QFQLRALMQKARKTAGLSDLY 380
              |||||||||||||||||||||
Db        361 QFQLRALMQKARKTAGLSDLY 381


RESULT 9
US-08-786-999-3
; Sequence 3, Application US/08786999
; Patent No. 5892016
;  GENERAL INFORMATION:
;    APPLICANT:  La Brie, Sam
;    APPLICANT:  Goli, Surya K.
;    TITLE OF INVENTION:  NOVEL HUMAN TUMOR
;    TITLE OF INVENTION:  SUPPRESSOR
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/786,999
;      FILING DATE:  Filed Herewith
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
```

```
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0199 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  381 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  GenBank
;      CLONE:  1330330
US-08-786-999-3

  Query Match             94.9%;  Score 1900.5;  DB 2;  Length 381;
  Best Local Similarity   94.5%;  Pred. No. 6.9e-147;
  Matches  360;  Conservative   11;  Mismatches    9;  Indels    1;  Gaps    1;

Qy          1 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 60
              |:||||||||||:|||| :||||||||||||||||:||||||| |||| |||
Db          1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNI 60

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120
              ||||||||||||||||||||||||||||||||||||||||:||||:|:|| |||:
Db         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQI 120

Qy        121 MIVVFGDEPPVFSRP-ISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCP 179
              |||:||:|||||||| :||||||| |||||||||||||| ||| ||||||||||||||||
Db        121 MIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCP 180

Qy        180 YPPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRA 239
              ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        181 YPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGA 240

Qy        240 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 300

Qy        300 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 360

Qy        360 QFQLRALMQKARKTAGLSDLY 380
              |||||||||||||||||||||
Db        361 QFQLRALMQKARKTAGLSDLY 381


RESULT 10
US-09-146-187-2
; Sequence 2, Application US/09146187
; Patent No. 6248523
;  GENERAL INFORMATION:
;    APPLICANT:  LI, Limin
;    APPLICANT:  COHEN, Stanley N
;    TITLE OF INVENTION:  MAMMALIAN TUMOR SUSCEPTIBILITY GENES AND
;    TITLE OF INVENTION:  THEIR USES
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FISH AND RICHARDSON, P.C.
;      STREET:  2200 SAND HILL ROAD
;      CITY:  MENLO PARK
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94025
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/146,187
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/670,274
;      FILING DATE:  June 13, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SHERWOOD, Pamela J.
;      REGISTRATION NUMBER:  36,677
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-781-1989
;      TELEFAX:  415-398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  381 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-146-187-2

  Query Match             94.9%;  Score 1900.5;  DB 4;  Length 381;
  Best Local Similarity   94.5%;  Pred. No. 6.9e-147;
  Matches  360;  Conservative   11;  Mismatches    9;  Indels    1;  Gaps    1;

Qy          1 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 60
              |:||||||||||:|||| :||||||||||||||||:||||||| |||| |||
Db          1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNI 60

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120
              ||||||||||||||||||||||||||||||||||||||||:||||:|:|| |||:
Db         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQI 120

Qy        121 MIVVFGDEPPVFSRP-ISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCP 179
              |||:||:|||||||| :||||||| |||||||||||||| ||| ||||||||||||||||
Db        121 MIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCP 180

Qy        180 YPPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRA 239
              ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        181 YPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGA 240

Qy        240 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 300

Qy        300 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 360

Qy        360 QFQLRALMQKARKTAGLSDLY 380
              |||||||||||||||||||||
Db        361 QFQLRALMQKARKTAGLSDLY 381


RESULT 11
US-08-999-774A-8
; Sequence 8, Application US/08999774A
; Patent No. 6274312
;  GENERAL INFORMATION:
;    APPLICANT:  Gish, Kurt C.
;    APPLICANT:  Seghezzi, Wolfgang
;    APPLICANT:  Shanahan, Frances
;    APPLICANT:  Lees, Emma M.
;    APPLICANT:  McClanahan, Terrill K.
;    TITLE OF INVENTION:  Intracellular Regulatory Molecules;
;    TITLE OF INVENTION:  Related Reagents
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DNAX Research Institute
;      STREET:  901 California Avenue
```

```
;       CITY:  Palo Alto
;       STATE:  California
;       COUNTRY:  USA
;       ZIP:  94304-1104
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/999,774A
;       FILING DATE:  10-DEC-1997
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 60/032,818
;       FILING DATE:  11-DEC-1996
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Ching, Edwin P.
;       REGISTRATION NUMBER:  34,090
;       REFERENCE/DOCKET NUMBER:  DX0646
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (650)852-9196
;       TELEFAX:  (650)496-1200
;   INFORMATION FOR SEQ ID NO:  8:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  237 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-999-774A-8

  Query Match             55.8%;  Score 1117;  DB 4;  Length 237;
  Best Local Similarity   98.0%;  Pred. No. 2e-83;
  Matches  201;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          11 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 70

Qy          61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          71 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 130

Qy         121 MIVVFGDEPPVFSRPISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCPY 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         131 MIVVFGDEPPVFSRPISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCPY 190

Qy         181 PPGGPYPATTSSQYPSQPPVTTVGP 205
              |||||||||||||||||||  : |
Db         191 PPGGPYPATTSSQYPSQPPVPLLVP 215


RESULT 12
US-08-526-136-13
; Sequence 13, Application US/08526136
; Patent No. 6107089
;  GENERAL INFORMATION:
;    APPLICANT:  Towle, Christine A. et al.
;    TITLE OF INVENTION:  ANNEXIN XI
;    NUMBER OF SEQUENCES:  36
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      COMPUTER:  IBM PS/2 Model 50Z or 55SX
;      OPERATING SYSTEM:  IBM P.C. DOS (Version 3.30)
;      SOFTWARE:  WordPerfect (Version 5.0)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/526,136
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/214,036
;      FILING DATE:
;      APPLICATION NUMBER:  07/837,775
;      FILING DATE:  February 13, 1992
;      APPLICATION NUMBER:  07/764,465
;      FILING DATE:  September 23, 1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Paul T.
;      REGISTRATION NUMBER:  30,162
;      REFERENCE/DOCKET NUMBER:  00786/099001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 542-5070
;      TELEFAX:  (617) 542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  466
;      TYPE:  amino acid
;      STRANDEDNESS:  N/A
;      TOPOLOGY:  N/A
US-08-526-136-13

  Query Match             8.0%;  Score 160;  DB 3;  Length 466;
  Best Local Similarity   21.6%;  Pred. No. 4e-05;
  Matches   84;  Conservative   37;  Mismatches   90;  Indels  178;  Gaps   17;

Qy         139 SYPPYQATG-PPNTSYMPG------MPGGISPYPSGYP---------------------- 169
               ||| |  || ||   | |         | |  |||||:|
Db           2 SYPGYPPTGYPPFPGYPPAGQESSFPPSGQYPYPSGFPPMGGGAYPQVPSSGYPGAGGYP 61

Qy         170 -----PNPSGYPGCPYPPGGP-YPATTSSQ-------------YPSQPPVTTVG------ 204
                  | | |||| | | | | ||     |              || |||  : |
Db          62 APGGYPAPGGYPGAPQPGGAPSYPGVPPGQGFGVPPGGAGFSGYP-QPPSQSYGGGPAQV 120

Qy         205 ----------------------PSRDGTISE------------DTIRASLI--------- 221
                                    ||:  |:::            | || : |
Db         121 PLPGGFPGGQMPSQYPGGQPTYPSQPATVTQVTQGTIRPAANFDAIRDAEILRKAMKGFG 180

Qy         222 ---SAVSDKLRWRMKEEMDRAQAELNAL--KRTEEDLKK---------------------- 255
                 |: | :  |  ::  : :|       |   :|||
Db         181 TDEQAIVDVVANRSNDQRQKIKAAFKTSYGKDLIKDLKSELSGNMEELILALFMPPTYYD 240

Qy         256 -----------GHQK---LEEMVTRLDQEVAEV------DKNIELLKKKDEELSSALEKM 295
                        | |:   :| : || :||: |:      :   :| |    : |   |:
Db         241 AWSLRKAMQGAGTQERVLIEILCTRTNQEIREIVRCYQSEFGRDLEKDIRSDTSGHFERL 300

Qy         296 ----------ENQSENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVID 345
                        |||| |: :                  |:|:|    ::  ||    | :
Db         301 LVSMCQGNRDENQSINHQM-----------------AQEDA--QRLYQAGE----GRLG 336

Qy         346 LDVFLKHVRLLSRKQFQLRALMQKARKTA 374
                |    :: | :|   |||| |:   : |
Db         337 TDESCFNMILATRSFPQLRATMEAYSRMA 365


RESULT 13
US-08-574-959A-9
; Sequence 9, Application US/08574959A
; Patent No. 5962224
;  GENERAL INFORMATION:
;    APPLICANT:  Jaekyoon Shin, Insil Joung, Ratna K. Vadlamudi
;    APPLICANT:  and Jack L. Strominger
;    TITLE OF INVENTION:  p62 POLYPEPTIDES, RELATED POLYPEPTIDES
;    TITLE OF INVENTION:  AND USES THEREFOR
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 State Street, Suite 510
;      CITY:  Boston
```

```
;       STATE:  Massachusetts
;       COUNTRY:  USA
;       ZIP:  02109-1875
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/574,959A
;       FILING DATE:  19-DEC-95
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Mandragouras, Amy E.
;       REGISTRATION NUMBER:  36,207
;       REFERENCE/DOCKET NUMBER:  DFN-008
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617)227-7400
;       TELEFAX:  (617)227-5941
;   INFORMATION FOR SEQ ID NO:  9:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  905 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-574-959A-9

  Query Match             7.0%;  Score 141;  DB 2;  Length 905;
  Best Local Similarity   23.4%;  Pred. No. 0.0036;
  Matches   68;  Conservative   45;  Mismatches  140;  Indels   38;  Gaps   10;

Qy         43 MNLTGTIP-VPYRGNTYNIPICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDA------ 95
              |   | :|  |:   | |:   | ||      ||     | :    : | : |
Db        457 MPSAGPVPSEPWTSTTANL---LGLLSRPSVCPPRLLPGPENH---RAGSNEDPILAPSG 510

Qy         96 ------------NGKIYLP-YLHEWKHPQSDLLGLIQ------VMIVVFGDEPPVFSRPI 136
                         |::  | ::|  |   ||:   ::      |:|| |  |     |
Db        511 TPPPTIPPDETFGGRVPRPAFVHYDKEEASDVEISLESDSDDSVVIVPEGLPPLPPPPPS 570

Qy        137 SASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCPYPPGGPYPATTSSQYPS 196
              |: ||   ||||  |  | :|      |   |  |   |  | || | |       |
Db        571 GATPPPIAPTGPPTAS--PPVPA--KEEPEELPAAPGPLPP-PPPPPPPVPGPVXLPPPQ 625

Qy        197 QPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRAQAELNALKRTEEDLKKG 256
                |  | |      : ||    :: |:  ::     :|| :  : |    :  ||: :
Db        626 LVPEGTPGGGGPPALEEDLTVININSSDEEEEEEGEEEEEEEEEEEEEEEEEEEEEDF 685

Qy        257 HQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQSENNDIDEV 307
              :: |:     ::|  | ::  |  :::: ||    |: |:: |  :::||
Db        686 EEEEEDEEEYFEEEEEEEEEFEEEFEEEEGELEEE-EEEEDEEEEEELEEV 735


RESULT 14
US-09-357-014-9
; Sequence 9, Application US/09357014
; Patent No. 6291645
;   GENERAL INFORMATION:
;       APPLICANT: Jaekyoon Shin, Insil Joung, Ratna K. Vadlamudi
;                  and Jack L. Strominger
;       TITLE OF INVENTION: p62 POLYPEPTIDES, RELATED POLYPEPTIDES
;                           AND USES THEREFOR
;       NUMBER OF SEQUENCES: 22
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: LAHIVE & COCKFIELD
;            STREET: 60 State Street, Suite 510
;            CITY: Boston
;            STATE: Massachusetts
;            COUNTRY: USA
;            ZIP: 02109-1875
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.25
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/357,014
;            FILING DATE: 19-Jul-1999
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/574,959
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Mandragouras, Amy E.
;            REGISTRATION NUMBER: 36,207
;            REFERENCE/DOCKET NUMBER: DFN-008
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (617)227-7400
;            TELEFAX: (617)227-5941
;   INFORMATION FOR SEQ ID NO: 9:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 905 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-09-357-014-9

  Query Match             7.0%;  Score 141;  DB 4;  Length 905;
  Best Local Similarity   23.4%;  Pred. No. 0.0036;
  Matches   68;  Conservative   45;  Mismatches  140;  Indels   38;  Gaps   10;

Qy         43 MNLTGTIP-VPYRGNTYNIPICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDA------ 95
              |   | :|  |:   | |:   | ||      ||     | :    : | : |
Db        457 MPSAGPVPSEPWTSTTANL---LGLLSRPSVCPPRLLPGPENH---RAGSNEDPILAPSG 510

Qy         96 ------------NGKIYLP-YLHEWKHPQSDLLGLIQ------VMIVVFGDEPPVFSRPI 136
                         |::  | ::|  |   ||:   ::      |:|| |  |     |
Db        511 TPPPTIPPDETFGGRVPRPAFVHYDKEEASDVEISLESDSDDSVVIVPEGLPPLPPPPPS 570

Qy        137 SASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCPYPPGGPYPATTSSQYPS 196
              |: ||   ||||  |  | :|      |   |  |   |  | || | |       |
Db        571 GATPPPIAPTGPPTAS--PPVPA--KEEPEELPAAPGPLPP-PPPPPPPVPGPVXLPPPQ 625

Qy        197 QPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRAQAELNALKRTEEDLKKG 256
                |  | |      : ||    :: |:  ::     :|| :  : |    :  ||: :
Db        626 LVPEGTPGGGGPPALEEDLTVININSSDEEEEEEGEEEEEEEEEEEEEEEEEEEEEDF 685

Qy        257 HQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQSENNDIDEV 307
              :: |:     ::|  | ::  |  :::: ||    |: |:: |  :::||
Db        686 EEEEEDEEEYFEEEEEEEEEFEEEFEEEEGELEEE-EEEEDEEEEEELEEV 735


RESULT 15
US-08-574-959A-7
; Sequence 7, Application US/08574959A
; Patent No. 5962224
;  GENERAL INFORMATION:
;    APPLICANT:  Jaekyoon Shin, Insil Joung, Ratna K. Vadlamudi
;    APPLICANT:  and Jack L. Strominger
;    TITLE OF INVENTION:  p62 POLYPEPTIDES, RELATED POLYPEPTIDES
;    TITLE OF INVENTION:  AND USES THEREFOR
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 State Street, Suite 510
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/574,959A
;      FILING DATE:  19-DEC-95
;    ATTORNEY/AGENT INFORMATION:
```

```
;       NAME:  Mandragouras, Amy E.
;       REGISTRATION NUMBER:  36,207
;       REFERENCE/DOCKET NUMBER:  DFN-008
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617)227-7400
;       TELEFAX:  (617)227-5941
;   INFORMATION FOR SEQ ID NO:  7:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1135 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-574-959A-7

  Query Match             7.0%;  Score 141;  DB 2;  Length 1135;
  Best Local Similarity   23.4%;  Pred. No. 0.0049;
  Matches   68;  Conservative   45;  Mismatches  140;  Indels   38;  Gaps   10;

Qy         43 MNLTGTIP-VPYRGNTYNIPICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDA------ 95
              |   | :|  |:   | |:   | ||      ||     | :    : | : |
Db        687 MPSAGPVPSEPWTSTTANL---LGLLSRPSVCPPRLLPGPENH---RAGSNEDPILAPSG 740

Qy         96 ------------NGKIYLP-YLHEWKHPQSDLLGLIQ------VMIVVFGDEPPVFSRPI 136
                          |:  | ::|  |   ||:   ::      |:||  |  |     |
Db        741 TPPPTIPPDETFGGRVPRPAFVHYDKEEASDVEISLESDSDDSVVIVPEGLPPLPPPPPS 800

Qy        137 SASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCPYPPGGPYPATTSSQYPS 196
              |: ||   ||||  |  | :|      |   |  |   |  | ||  | |       |
Db        801 GATPPPIAPTGPPTAS--PPVPA--KEEPEELPAAPGPLPP-PPPPPPPVPGPVXLPPPQ 855

Qy        197 QPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRAQAELNALKRTEEDLKKG 256
              |  | |     : ||    :: |:  ::     :|| :  : |    : ||: :
Db        856 LVPEGTPGGGGPPALEEDLTVININSSDEEEEEEGEEEEEEEEEEEEEEEEEEEEEEDF 915

Qy        257 HQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQSENNDIDEV 307
              :: |:     ::|  | ::  |  :::: ||    |: |:: |  :::||
Db        916 EEEEEDEEEYFEEEEEEEEFEEEFEEEEGELEEE-EEEEDEEEEELEEV 965


Search completed: June  3, 2003, 14:08:57
Job time : 13.9829 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: June 3, 2003, 14:09:05 ; Search time 30.4599 Seconds
(without alignments)
1262.807 Million cell updates/sec

Title: US-09-804-690-4
Perfect score: 2002
Sequence: 1 MVSKYKYRDLTVRETVNVIT.....FQLRALMQKARKTAGLSDLY 380

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 383519 seqs, 101223694 residues

Total number of hits satisfying chosen parameters: 383519

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2002	100.0	380	10	US-09-804-690-4	Sequence 4, Appli
2	1900.5	94.9	381	10	US-09-804-690-2	Sequence 2, Appli
3	346	17.3	146	9	US-09-925-299-1257	Sequence 1257, Ap
4	346	17.3	146	10	US-09-925-299-1257	Sequence 1257, Ap
5	200	10.0	296	10	US-09-801-368-386	Sequence 386, App
6	161	8.0	485	10	US-09-925-300-1664	Sequence 1664, Ap
7	148.5	7.4	148	9	US-10-012-542-453	Sequence 453, App
8	132	6.6	371	9	US-09-284-320-2	Sequence 2, Appli
9	132	6.6	371	9	US-10-197-666A-4	Sequence 4, Appli
10	130	6.5	1647	9	US-09-824-574-4	Sequence 4, Appli
11	127.5	6.4	284	9	US-10-063-547-62	Sequence 62, Appl
12	127.5	6.4	284	9	US-10-174-590-236	Sequence 236, App
13	127.5	6.4	284	9	US-10-176-758-236	Sequence 236, App
14	127.5	6.4	284	9	US-10-063-616-62	Sequence 62, Appl
15	127.5	6.4	284	9	US-10-175-737-236	Sequence 236, App
16	127.5	6.4	284	9	US-10-063-502-62	Sequence 62, Appl
17	127.5	6.4	284	9	US-10-173-706-236	Sequence 236, App
18	127.5	6.4	284	9	US-10-175-738-236	Sequence 236, App
19	127.5	6.4	284	9	US-10-175-752-236	Sequence 236, App
20	127.5	6.4	284	9	US-10-176-482-236	Sequence 236, App
21	127.5	6.4	284	9	US-10-176-757-236	Sequence 236, App
22	127.5	6.4	284	9	US-10-176-913-236	Sequence 236, App
23	127.5	6.4	284	9	US-10-180-552-236	Sequence 236, App
24	127.5	6.4	284	9	US-10-180-557-236	Sequence 236, App
25	127.5	6.4	284	9	US-10-173-700-236	Sequence 236, App
26	127.5	6.4	284	9	US-10-174-572-236	Sequence 236, App
27	127.5	6.4	284	9	US-10-174-579-236	Sequence 236, App
28	127.5	6.4	284	9	US-10-174-582-236	Sequence 236, App
29	127.5	6.4	284	9	US-10-174-588-236	Sequence 236, App
30	127.5	6.4	284	9	US-10-175-739-236	Sequence 236, App
31	127.5	6.4	284	9	US-10-175-740-236	Sequence 236, App
32	127.5	6.4	284	9	US-10-175-743-236	Sequence 236, App
33	127.5	6.4	284	9	US-10-176-488-236	Sequence 236, App
34	127.5	6.4	284	9	US-10-176-492-236	Sequence 236, App
35	127.5	6.4	284	9	US-10-176-747-236	Sequence 236, App
36	127.5	6.4	284	9	US-10-176-750-236	Sequence 236, App
37	127.5	6.4	284	9	US-10-176-985-236	Sequence 236, App
38	127.5	6.4	284	9	US-10-176-987-236	Sequence 236, App
39	127.5	6.4	284	9	US-10-176-991-236	Sequence 236, App
40	127.5	6.4	284	9	US-10-176-992-236	Sequence 236, App
41	127.5	6.4	284	9	US-10-176-993-236	Sequence 236, App
42	127.5	6.4	284	9	US-10-184-658-236	Sequence 236, App
43	127.5	6.4	284	9	US-10-173-695-236	Sequence 236, App
44	127.5	6.4	284	9	US-10-173-697-236	Sequence 236, App
45	127.5	6.4	284	9	US-10-173-705-236	Sequence 236, App

ALIGNMENTS

```
RESULT 1
US-09-804-690-4
; Sequence 4, Application US/09804690
; Patent No. US20020034743A1
;  GENERAL INFORMATION:
;       APPLICANT: LI, Limin
;                  COHEN, Stanley N
;       TITLE OF INVENTION: MAMMALIAN TUMOR SUSCEPTIBILITY GENES AND
;                           THEIR USES
;       NUMBER OF SEQUENCES: 20
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: FISH AND RICHARDSON, P.C.
;            STREET: 2200 SAND HILL ROAD
;            CITY: MENLO PARK
;            STATE: CA
;            COUNTRY: USA
;            ZIP: 94025
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/804,690
;            FILING DATE: 12-Mar-2001
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 09/146,187
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: SHERWOOD, Pamela J.
;            REGISTRATION NUMBER: 36,677
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 415-781-1989
;            TELEFAX: 415-398-3249
;            TELEX: 910 277299
;  INFORMATION FOR SEQ ID NO: 4:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 380 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
```

```
;       SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-804-690-4

  Query Match            100.0%;  Score 2002;  DB 10;  Length 380;
  Best Local Similarity  100.0%;  Pred. No. 3.7e-133;
  Matches  380;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 60

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120

Qy        121 MIVVFGDEPPVFSRPISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCPY 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 MIVVFGDEPPVFSRPISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCPY 180

Qy        181 PPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRAQ 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 PPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRAQ 240

Qy        241 AELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQSE 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQSE 300

Qy        301 NNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLLDVFLKHVRLLSRKQ 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 NNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLLDVFLKHVRLLSRKQ 360

Qy        361 FQLRALMQKARKTAGLSDLY 380
              ||||||||||||||||||||
Db        361 FQLRALMQKARKTAGLSDLY 380


RESULT 2
US-09-804-690-2
; Sequence 2, Application US/09804690
; Patent No. US20020034743A1
;   GENERAL INFORMATION:
;       APPLICANT: LI, Limin
;                  COHEN, Stanley N
;       TITLE OF INVENTION: MAMMALIAN TUMOR SUSCEPTIBILITY GENES AND
;                           THEIR USES
;       NUMBER OF SEQUENCES: 20
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: FISH AND RICHARDSON, P.C.
;            STREET: 2200 SAND HILL ROAD
;            CITY: MENLO PARK
;            STATE: CA
;            COUNTRY: USA
;            ZIP: 94025
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/804,690
;            FILING DATE: 12-Mar-2001
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 09/146,187
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: SHERWOOD, Pamela J.
;            REGISTRATION NUMBER: 36,677
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 415-781-1989
;            TELEFAX: 415-398-3249
;            TELEX: 910 277299
;   INFORMATION FOR SEQ ID NO: 2:
```

```
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 381 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-804-690-2

  Query Match            94.9%;  Score 1900.5;  DB 10;  Length 381;
  Best Local Similarity  94.5%;  Pred. No. 5.1e-126;
  Matches  360;  Conservative   11;  Mismatches    9;  Indels    1;  Gaps    1;

Qy          1 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 60
              |:|||||||||||:||||| :|||||||||||||||||||:|||||||| |||| |||
Db          1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNI 60

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120
              ||||||||||||||||||||||||||||||||||||||||||:||||:|:|| |||:
Db         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQI 120

Qy        121 MIVVFGDEPPVFSRP-ISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCP 179
              |||:||:|||||||| :||||||| |||||||||||||| ||| ||||||||||||||
Db        121 MIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCP 180

Qy        180 YPPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRA 239
              ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        181 YPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGA 240

Qy        240 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 300

Qy        300 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 360

Qy        360 QFQLRALMQKARKTAGLSDLY 380
              ||||||||||||||||||||
Db        361 QFQLRALMQKARKTAGLSDLY 381


RESULT 3
US-09-925-299-1257
; Sequence 1257, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA102
;  CURRENT APPLICATION NUMBER: US/09/925,299
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05883
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 1556
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1257
;   LENGTH: 146
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (2)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (6)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (12)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (131)
```

```
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (135)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (138)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (145)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-1257

  Query Match             17.3%;  Score 346;  DB 9;  Length 146;
  Best Local Similarity   95.5%;  Pred. No. 3.1e-17;
  Matches   63;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy         33 VFNDGSSRELMNLTGTIPVPYRGNTYNIPICLWLLDTYPYNPPICFVKPTSSMTIKTGKH 92
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         69 VFNDGSSRELMNLTGTIPVPYRGNTYNIPICLWLLDTYPYNPPICFVKPTSSMTIKTGKH 128

Qy         93 VDANGK 98
              ||   |
Db        129 VDXPKK 134


RESULT 4
US-09-925-299-1257
; Sequence 1257, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA102
;  CURRENT APPLICATION NUMBER: US/09/925,299
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05883
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 1556
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1257
;   LENGTH: 146
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (2)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (6)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (12)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (131)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (135)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (138)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (145)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-1257

  Query Match             17.3%;  Score 346;  DB 10;  Length 146;
  Best Local Similarity   95.5%;  Pred. No. 3.1e-17;
  Matches   63;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy         33 VFNDGSSRELMNLTGTIPVPYRGNTYNIPICLWLLDTYPYNPPICFVKPTSSMTIKTGKH 92
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         69 VFNDGSSRELMNLTGTIPVPYRGNTYNIPICLWLLDTYPYNPPICFVKPTSSMTIKTGKH 128

Qy         93 VDANGK 98
              ||   |
Db        129 VDXPKK 134


RESULT 5
US-09-801-368-386
; Sequence 386, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
;  APPLICANT: Busby, Robert
;  APPLICANT:  Cali, Brian
;  APPLICANT:  Hecht, Peter
;  APPLICANT:  Holtzman, Doug
;  APPLICANT:  Madden, Kevin
;  APPLICANT:  Maxon, Mary
;  APPLICANT:  Milne, Todd
;  APPLICANT:  No. US20020128250A1man, Thea
;  APPLICANT:  Royer, John
;  APPLICANT:  Salama, Sofie
;  APPLICANT:  Sherman, Amir
;  APPLICANT:  Silva, Jeff
;  APPLICANT:  Summers, Eric
;  TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
;  FILE REFERENCE: 109272.147
;  CURRENT APPLICATION NUMBER: US/09/801,368
;  CURRENT FILING DATE:  2001-03-07
;  PRIOR APPLICATION NUMBER: US 09/487,558
;  PRIOR FILING DATE: 2000-01-19
;  PRIOR APPLICATION NUMBER: US 60/160,587
;  PRIOR FILING DATE: 1999-10-20
;  NUMBER OF SEQ ID NOS: 440
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 386
;   LENGTH: 296
;   TYPE: PRT
;   ORGANISM: Saccharomyces cerevisiae
US-09-801-368-386

  Query Match             10.0%;  Score 200;  DB 10;  Length 296;
  Best Local Similarity   24.2%;  Pred. No. 1.4e-06;
  Matches   72;  Conservative   57;  Mismatches  124;  Indels   44;  Gaps   13;

Qy          7 YRD--LTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNT-YNIPIC 63
              | |   |  ::: ::  :  |:|    :  :||: : |::: |||     |:: ::||:
Db         25 YNDGRTTFHDSLALLDNFHSLRPRTRVFTHSDGTPQLLLSIYGTISTGEDGSSPHSIPVI 84

Qy         64 LWLLDTYPYNPPIC------FVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGL 117
              :|:   ||  ||        |   | | ::   :::|:|| | || || |     :|: :
Db         85 MWVPSMYPVKPPFISINLENFDMNTISSSLPIQEYIDSNGWIALPILHCWDPAAMNLIMV 144

Qy        118 IQVMIVVFGDEPPVFSRPISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPG 177
              :| :: :   |||    |     | |     ||  :|  |     |  || |
Db        145 VQELMSLL-HEPPQDQAPSLPPKPNTQLQQEQNTPPLPPKPKS----PHLKPPLP----- 194

Qy        178 CPYPPGGPYPATTSSQYPSQPPVTTVGPSRD-------GTISEDTIRASLISAVSDKLRW 230
              ||  | ||: :          | : |:          |:  :  |   :  |:||:
Db        195 ---PPPPPQPASNALDLMDMDN-TDISPTNHHEMLQNLQTVVNELYRED-VDYVADKILT 249

Qy        231 R---MKEEMDRAQAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKK 284
              |   |:| : |   |: |       : | | :  |    ::| :  ::  |::| :|
Db        250 RQTVMQESIARFH-EIIA-------IDKNHLRAVEQA--IEQTMHSLNAQIDVLNRK 296


RESULT 6
US-09-925-300-1664
; Sequence 1664, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
```

```
;  APPLICANT: Craig Rosen,
;  APPLICANT:  Steve Ruben
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA101
;  CURRENT APPLICATION NUMBER: US/09/925,300
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05988
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 1890
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1664
;   LENGTH: 485
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-925-300-1664

  Query Match             8.0%;  Score 161;  DB 10;  Length 485;
  Best Local Similarity   21.4%;  Pred. No. 0.0015;
  Matches   84;  Conservative   38;  Mismatches   92;  Indels  178;  Gaps   17;

Qy        136 ISASYPPYQATG-PPNTSYMPG------MPGGISPYPSGYP------------------- 169
              :  ||| |  || ||   | |        | |  |||||:|
Db         18 VRMSYPGYPPTGYPPFPGYPPAGQESSFPPSGQYPYPSGFPPMGGGAYPQVPSSGYPGAG 77

Qy        170 --------PNPSGYPGCPYPPGGP-YPATTSSQ-------------YPSQPPVTTVG--- 204
                      | | |||| | | | | ||     |              || |||  : |
Db         78 GYPAPGGYPAPGGYPGAPQPGGAPSYPGVPPGQGFGVPPGGAGFSGYP-QPPSQSYGGGP 136

Qy        205 -------------------------PSRDGTISE-----------DTIRASLI------ 221
                                       ||:  |:::           | || : |
Db        137 AQVPLPGGFPGGQMPSQYPGGQPTYPSQPATVTQVTQGTIRPAANFDAIRDAEILRKAMK 196

Qy        222 ------SAVSDKLRWRMKEEMDRAQAELNAL--KRTEEDLKK------------------ 255
                   |: | :  |  ::  : :|       |   :|||
Db        197 GFGTDEQAIVDVVANRSNDQRQKIKAAFKTSYGKDLIKDLKSELSGNMEELILALFMPPT 256

Qy        256 --------------GHQK---LEEMVTRLDQEVAEV------DKNIELLKKKDEELSSAL 292
                         | |:   :| : || :||: |:      :  :| |    : |
Db        257 YYDAWSLRKAMQGAGTQERVLIEILCTRTNQEIREIVRCYQSEFGRDLEKDIRSDTSGHF 316

Qy        293 EKM----------ENQSENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRG 342
              |::          |||| |: :                  |:|:|    ::  ||    | .
Db        317 ERLLVSMCQGNRDENQSINHQM-----------------AQEDA--QRLYQAGE----G 352

Qy        343 VIDLDVFLKHVRLLSRKQFQLRALMQKARKTA 374
               :  |    :: | :|   ||||  |:   : |
Db        353 RLGTDESCFNMILATRSFPQLRATMEAYSRMA 384


RESULT 7
US-10-012-542-453
; Sequence 453, Application US/10012542
; Publication No. US20030044851A1
; GENERAL INFORMATION:
;  APPLICANT: Ruben et al.
;  TITLE OF INVENTION: 94 Human Secreted Proteins
;  FILE REFERENCE: PZ029P1
;  CURRENT APPLICATION NUMBER: US/10/012,542
;  CURRENT FILING DATE:  2001-12-12
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/461,325
;  PRIOR FILING DATE: EARLIER FILING DATE: 1999-12-14
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,507
;  PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,508
;  PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,509
;  PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,510
;  PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/090,112
;  PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-22
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/090,113
;  PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-22
;  NUMBER OF SEQ ID NOS: 532
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 453
;   LENGTH: 148
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-012-542-453

  Query Match             7.4%;  Score 148.5;  DB 9;  Length 148;
  Best Local Similarity   29.7%;  Pred. No. 0.0024;
  Matches   41;  Conservative   18;  Mismatches   46;  Indels   33;  Gaps    6;

Qy        126 GDEPPVFSRPISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNP---SGYPGC---- 178
              |   |    ||:  :||      |       || | |   :| | ||:|    ||||
Db         22 GGSNPAHPPPINPPFPPGPCPPP------PGAPHGNPAFPPGGPPHPVPQPGYPGCQPLG 75

Qy        179 PYPPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDR 238
              ||||  | ||      |  |||  : |   |                 : : ||   :|:::| :
Db         76 PYPPPYPPPA------PGIPPVNPLAPGMVGP----------AVIVDK---KMQKKMKK 115

Qy        239 AQAELNALKRTEEDLKKG 256
              |  :::  ::  :  | |
Db        116 AHKKMHKHQKHHKYHKHG 133


RESULT 8
US-09-284-320-2
; Sequence 2, Application US/09284320
; Publication No. US20030092175A1
; GENERAL INFORMATION:
;  APPLICANT:  Kato, Seishi et al.
;  TITLE OF INVENTION:  HUMAN PROTEINS HAVING TRANSMEMBRANE DOMAINS AND DNAs
;  TITLE OF INVENTION:  ENCODING THESE PROTEINS
;  FILE REFERENCE:  GIN-6705CPUS
;  CURRENT APPLICATION NUMBER: US/09/284,320
;  CURRENT FILING DATE:  1999-06-21
;  PRIOR APPLICATION NUMBER:  JP 8-301429
;  PRIOR FILING DATE:  1996-11-13
;  PRIOR APPLICATION NUMBER:  PCT/JP97/04056
;  PRIOR FILING DATE:  1997-11-07
;  NUMBER OF SEQ ID NOS:  91
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 371
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-284-320-2

  Query Match             6.6%;  Score 132;  DB 9;  Length 371;
  Best Local Similarity   41.6%;  Pred. No. 0.12;
  Matches   37;  Conservative    7;  Mismatches   25;  Indels   20;  Gaps    8;

Qy        126 GDEPPV--FSRP--ISASY--PPYQATGPPNTSYMPGMPGGISPYP-SGYPPNP---SGY 175
              | :||:  :::|    | |  ||:|    |:    || | | ||||  |||  |    ||
Db         24 GPQPPMPPYAQPPYPGAPYPQPPFQ----PSPYGQPGYPHGPSPYPQGGYPQGPYPQGGY 79

Qy        176 PGCPYP----PGGPYPATTSSQ--YPSQP 198
              |  |||    | ||||     |  ||  |
Db         80 PQGPYPQEGYPQGPYPQGGYPQGPYPQSP 108


RESULT 9
US-10-197-666A-4
; Sequence 4, Application US/10197666A
; Publication No. US20030092037A1
; GENERAL INFORMATION:
;  APPLICANT: ASAHI KASEI KABUSIKI KAISYA
;  TITLE OF INVENTION: Elk1 phosphorylation related gene
;  FILE REFERENCE: PH-1548US
;  CURRENT APPLICATION NUMBER: US/10/197,666A
;  CURRENT FILING DATE:  2002-11-18
```

```
;  PRIOR APPLICATION NUMBER: JP 2001-218204
;  PRIOR FILING DATE: 2001-07-18
;  PRIOR APPLICATION NUMBER: JP 2001-263450
;  PRIOR FILING DATE: 2001-08-31
;  PRIOR APPLICATION NUMBER: JP 2002-012176
;  PRIOR FILING DATE: 2002-01-21
;  PRIOR APPLICATION NUMBER: US 60/305,884
;  PRIOR FILING DATE: 2001-07-18
;  PRIOR APPLICATION NUMBER: US 60/316,304
;  PRIOR FILING DATE: 2001-09-04
;  PRIOR APPLICATION NUMBER: US 60/350,027
;  PRIOR FILING DATE: 2002-01-23
;  NUMBER OF SEQ ID NOS: 156
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 371
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-197-666A-4

  Query Match             6.6%;  Score 132;  DB 9;  Length 371;
  Best Local Similarity   41.6%;  Pred. No. 0.12;
  Matches   37;  Conservative    7;  Mismatches   25;  Indels   20;  Gaps    8;

Qy        126 GDEPPV--FSRP--ISASY--PPYQATGPPNTSYMPGMPGGISPYP-SGYPPNP---SGY 175
              | :||:  :::|    | |  ||:|    |:    || | | |||| |||  |    ||
Db         24 GPQPPMPPYAQPPYPGAPYPQPPFQ----PSPYGQPGYPHGPSPYPQGGYPQGPYPQGGY 79

Qy        176 PGCPYP----PGGPYPATTSSQ--YPSQP 198
              |  |||    | ||||     |  ||  |
Db         80 PQGPYPQEGYPQGPYPQGGYPQGPYPQSP 108


RESULT 10
US-09-824-574-4
; Sequence 4, Application US/09824574
; Publication No. US20030077800A1
; GENERAL INFORMATION:
;  APPLICANT: Rouleau, Natalie
;  APPLICANT:  Moilanen, Anu-Maarit
;  APPLICANT:  Palvimo, Jorma J.
;  APPLICANT:  Jnne, Olli A,
;  TITLE OF INVENTION: ARIP4 Gene and Protein
;  FILE REFERENCE: 2630-109
;  CURRENT APPLICATION NUMBER: US/09/824,574
;  CURRENT FILING DATE:  2001-04-03
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
;   LENGTH: 1647
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-824-574-4

  Query Match             6.5%;  Score 130;  DB 9;  Length 1647;
  Best Local Similarity   21.7%;  Pred. No. 1.1;
  Matches   61;  Conservative   27;  Mismatches   71;  Indels  122;  Gaps   11;

Qy        126 GDEPPVFSRPISASYPPYQATGPPNTSYMPGMP-----GGISPYPSG------------ 167
              |  || :|||     |     ||  :   ||||     |   |:| |
Db        243 GPAPPNYSRPHGMGGPNMPPPGP--SGVPPGMPGQPPGGPPKPWPEGPMANAAAPTSTPQ 300

Qy        168 --YPPNPSGYPGCPYPPGGP-------YPATTSSQYPSQPP-----------VTTVGPSR 207
                 || |:| |  | ||  |        | | |   |:||            :| :   |
Db        301 KLIPPQPTGRPS-PAPPAVPPAASPVMPPQTQSPGQPAQPAPMVPLHQKQSRITPIQKPR 359

Qy        208 ----------------------------DGTISED-----TI---------------- 216
                                         |::: |     ||
Db        360 GLDPVEILQEREYRLQARIAHRIQELENLPGSLAGDLRTKATIELKALRLLNFQRQLRQE 419

Qy        217 ------------------------RASLISA-VSDKLRWRMKEEMDRA-----QAELNAL 246
                                   | ||  | :::||  : | | :|      |  ||::
Db        420 VVVCMRRDTALETALNAKAYKRSKRQSLREARITEKLEKQQKIEQERKRRQKHQEYLNSI 479

Qy        247 KRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEE 287
                :  :| |: |: :   : :| : ||    | |  :||: |
Db        480 LQHAKDFKEYHRSVTGKIQKLTKAVATYHANTEREQKKENE 520


RESULT 11
US-10-063-547-62
; Sequence 62, Application US/10063547
; Publication No. US20020182638A1
; GENERAL INFORMATION:
;  APPLICANT: Eaton,Dan L.
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Grimaldi,Christopher J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3230R1C1
;  CURRENT APPLICATION NUMBER: US/10/063,547
;  CURRENT FILING DATE:  2002-05-02
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 62
;   LENGTH: 284
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-063-547-62

  Query Match             6.4%;  Score 127.5;  DB 9;  Length 284;
  Best Local Similarity   23.6%;  Pred. No. 0.17;
  Matches   51;  Conservative   22;  Mismatches   68;  Indels   75;  Gaps    7;

Qy        138 ASYP--------PYQATGPPNTSYMPGMPGGISPYPSGYPPN----------------- 171
              ||||          || | |  || || |     | || ||
Db          2 ASYPYRQGCPGAAGQAPGAPPGSYYPGPPNSGGQYGSGLPPGGGYGGPAPGGPYGPPAGG 61

Qy        172 ----------PSGYPGCPY---PPGGPYPATTSSQYPSQPP----------------- 199
                        ||| || ||    |||||     | | :| |
Db         62 GPYGHPNPGMFPSGTPGGPYGGAAPGGPYGQPPPSSYGAQQPGLYGQGGAPPNVDPEAYS 121

Qy        200 -VTTVGPSRDGTISEDTIRASLI----SAVSDKLRWRMKEEMDRAQA------ELNALKR 248
                :|     | ||   :: :|:    |: :|:    |    |: ::        :|| :
Db        122 WFQSVDSDHSGYISMKELKQALVNCNWSSFNDETCLMMINMFDKTKSGRIDVYGFSALWK 181

Qy        249 TEEDLKKGHQKLEE------MVTRLDQEVAEVDKNI 278
                :  |   |: :         | | | ::::  |:
Db        182 FIQQWKNLFQQYDRDRSGSISYTELQQALSQMGYNL 217


RESULT 12
US-10-174-590-236
; Sequence 236, Application US/10174590
; Publication No. US20030008352A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Chen,Jian
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Pan,James
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  APPLICANT:  Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3430R1C42
```

```
;  CURRENT APPLICATION NUMBER: US/10/174,590
;  CURRENT FILING DATE:  2002-06-18
;  Prior application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 236
;   LENGTH: 284
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-174-590-236

  Query Match             6.4%;  Score 127.5;  DB 9;  Length 284;
  Best Local Similarity   23.6%;  Pred. No. 0.17;
  Matches   51;  Conservative   22;  Mismatches   68;  Indels   75;  Gaps    7;

Qy        138 ASYP--------PYQATGPPNTSYMPGMPGGISPYPSGYPPN----------------- 171
              ||||          || | |  || || |     | || ||
Db          2 ASYPYRQGCPGAAGQAPGAPPGSYYPGPPNSGGQYGSGLPPGGGYGGPAPGGPYGPPAGG 61

Qy        172 ----------PSGYPGCPY---PPGGPYPATTSSQYPSQPP----------------- 199
                      ||| || ||    |||||     | | :| |
Db         62 GPYGHPNPGMFPSGTPGGPYGGAAPGGPYGQPPPSSYGAQQPGLYGQGGAPPNVDPEAYS 121

Qy        200 -VTTVGPSRDGTISEDTIRASLI----SAVSDKLRWRMKEEMDRAQA------ELNALKR 248
               :|     | ||   :: :|:    |: :|:    |    |: ::        :|| : .
Db        122 WFQSVDSDHSGYISMKELKQALVNCNWSSFNDETCLMMINMFDKTKSGRIDVYGFSALWK 181

Qy        249 TEEDLKKGHQKLEE------MVTRLDQEVAEVDKNI 278
               :  |   |: :         | | | ::::  |:
Db        182 FIQQWKNLFQQYDRDRSGSISYTELQQALSQMGYNL 217


RESULT 13
US-10-176-758-236
; Sequence 236, Application US/10176758
; Publication No. US20030008353A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Chen,Jian
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Pan,James
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  APPLICANT:  Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3430R1C104
;  CURRENT APPLICATION NUMBER: US/10/176,758
;  CURRENT FILING DATE:  2002-06-21
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 236
;   LENGTH: 284
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-176-758-236

  Query Match             6.4%;  Score 127.5;  DB 9;  Length 284;
  Best Local Similarity   23.6%;  Pred. No. 0.17;
  Matches   51;  Conservative   22;  Mismatches   68;  Indels   75;  Gaps    7;

Qy        138 ASYP--------PYQATGPPNTSYMPGMPGGISPYPSGYPPN----------------- 171
              ||||          || | |  || || |     | || ||
Db          2 ASYPYRQGCPGAAGQAPGAPPGSYYPGPPNSGGQYGSGLPPGGGYGGPAPGGPYGPPAGG 61

Qy        172 ----------PSGYPGCPY---PPGGPYPATTSSQYPSQPP----------------- 199
                      ||| || ||    |||||     | | :| |
Db         62 GPYGHPNPGMFPSGTPGGPYGGAAPGGPYGQPPPSSYGAQQPGLYGQGGAPPNVDPEAYS 121

Qy        200 -VTTVGPSRDGTISEDTIRASLI----SAVSDKLRWRMKEEMDRAQA------ELNALKR 248
               :|     | ||   :: :|:    |: :|:    |    |: ::        :|| :
Db        122 WFQSVDSDHSGYISMKELKQALVNCNWSSFNDETCLMMINMFDKTKSGRIDVYGFSALWK 181

Qy        249 TEEDLKKGHQKLEE------MVTRLDQEVAEVDKNI 278
               :  |   |: :         | | | ::::  |:
Db        182 FIQQWKNLFQQYDRDRSGSISYTELQQALSQMGYNL 217


RESULT 14
US-10-063-616-62
; Sequence 62, Application US/10063616
; Publication No. US20030013855A1
; GENERAL INFORMATION:
;  APPLICANT: Eaton,Dan L.
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Grimaldi,Christopher J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3230R1C1
;  CURRENT APPLICATION NUMBER: US/10/063,616
;  CURRENT FILING DATE:  2002-05-03
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 62
;   LENGTH: 284
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-063-616-62

  Query Match             6.4%;  Score 127.5;  DB 9;  Length 284;
  Best Local Similarity   23.6%;  Pred. No. 0.17;
  Matches   51;  Conservative   22;  Mismatches   68;  Indels   75;  Gaps    7;

Qy        138 ASYP--------PYQATGPPNTSYMPGMPGGISPYPSGYPPN----------------- 171
              ||||          || | |  || || |     | || ||
Db          2 ASYPYRQGCPGAAGQAPGAPPGSYYPGPPNSGGQYGSGLPPGGGYGGPAPGGPYGPPAGG 61

Qy        172 ----------PSGYPGCPY---PPGGPYPATTSSQYPSQPP----------------- 199
                      ||| || ||    |||||     | | :| |
Db         62 GPYGHPNPGMFPSGTPGGPYGGAAPGGPYGQPPPSSYGAQQPGLYGQGGAPPNVDPEAYS 121

Qy        200 -VTTVGPSRDGTISEDTIRASLI----SAVSDKLRWRMKEEMDRAQA------ELNALKR 248
               :|     | ||   :: :|:    |: :|:    |    |: ::        :|| :
Db        122 WFQSVDSDHSGYISMKELKQALVNCNWSSFNDETCLMMINMFDKTKSGRIDVYGFSALWK 181

Qy        249 TEEDLKKGHQKLEE------MVTRLDQEVAEVDKNI 278
               :  |   |: :         | | | ::::  |:
Db        182 FIQQWKNLFQQYDRDRSGSISYTELQQALSQMGYNL 217


RESULT 15
US-10-175-737-236
; Sequence 236, Application US/10175737
; Publication No. US20030013153A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Chen,Jian
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Pan,James
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  APPLICANT:  Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
```

```
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3430R1C50
;  CURRENT APPLICATION NUMBER: US/10/175,737
;  CURRENT FILING DATE:  2002-06-19
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 236
;   LENGTH: 284
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-175-737-236

  Query Match             6.4%;  Score 127.5;  DB 9;  Length 284;
  Best Local Similarity   23.6%;  Pred. No. 0.17;
  Matches   51;  Conservative   22;  Mismatches   68;  Indels   75;  Gaps    7;

Qy        138 ASYP--------PYQATGPPNTSYMPGMPGGISPYPSGYPPN----------------- 171
              ||||          || | |  || || |     | || ||
Db          2 ASYPYRQGCPGAAGQAPGAPPGSYYPGPPNSGGQYGSGLPPGGGYGGPAPGGPYGPPAGG 61

Qy        172 -----------PSGYPGCPY---PPGGPYPATTSSQYPSQPP----------------- 199
                         ||| || ||    |||||     | | :| |
Db         62 GPYGHPNPGMFPSGTPGGPYGGAAPGGPYGQPPPSSYGAQQPGLYGQGGAPPNVDPEAYS 121

Qy        200 -VTTVGPSRDGTISEDTIRASLI----SAVSDKLRWRMKEEMDRAQA------ELNALKR 248
               :|     | ||   :: :|:    |: :|:    |    |: ::        :|| :
Db        122 WFQSVDSDHSGYISMKELKQALVNCNWSSFNDETCLMMINMFDKTKSGRIDVYGFSALWK 181

Qy        249 TEEDLKKGHQKLEE------MVTRLDQEVAEVDKNI 278
                :  |   |: :         | | | ::::  |:
Db        182 FIQQWKNLFQQYDRDRSGSISYTELQQALSQMGYNL 217


Search completed: June  3, 2003, 14:22:35
Job time : 31.4599 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: June 3, 2003, 14:06:00 ; Search time 25.9658 Seconds
(without alignments)
1406.890 Million cell updates/sec

Title: US-09-804-690-4
Perfect score: 2002
Sequence: 1 MVSKYKYRDLTVRETVNVIT.....FQLRALMQKARKTAGLSDLY 380

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	658	32.9	404	2	H88650	protein C09G12.9 [
2	388	19.4	83	2	I48283	gene CC2 protein -
3	153.5	7.7	169	2	T34520	hypothetical prote
4	147	7.3	488	1	LUHU7	annexin VII, long
5	145.5	7.3	212	2	S74288	hypothetical prote
6	142	7.1	505	2	A53152	annexin XI - human
7	141	7.0	463	2	S29170	annexin VII - mous
8	138.5	6.9	827	2	T39608	zinc finger transc
9	136.5	6.8	503	1	LURB11	annexin XI - rabbi
10	135.5	6.8	437	2	T14192	extensin homolog T
11	134	6.7	198	2	D70509	hypothetical prote
12	134	6.7	485	2	T37550	hypothetical coile
13	132.5	6.6	671	2	T36037	probable export as
14	131.5	6.6	370	2	T42532	hypothetical prote
15	131	6.5	1613	2	S39059	protein BRG1 - hum
16	131	6.5	1880	2	T18531	tractin - medicina
17	130.5	6.5	338	2	I53043	transforming prote
18	130	6.5	1647	2	S45252	SNF2beta protein -
19	128.5	6.4	1006	2	T42731	atrophin-1 related
20	128	6.4	139	2	H84809	hypothetical prote
21	128	6.4	3942	2	T42730	Bassoon protein -
22	127	6.3	1184	2	G01763	atrophin-1 - human
23	126.5	6.3	678	2	H88187	protein C18H9.8 [i
24	126	6.3	1453	2	S21626	collagen alpha 1(I
25	125	6.2	2715	2	T13049	eyelid - fruit fly
26	124.5	6.2	669	2	T28754	hypothetical prote
27	124.5	6.2	990	2	T14756	hypothetical prote
28	124	6.2	964	2	T21865	hypothetical prote
29	123.5	6.2	279	2	T05421	hypothetical prote
30	123.5	6.2	380	2	S51797	vasodilator-stimul
31	123	6.1	564	2	H70804	hypothetical prote
32	122	6.1	887	1	S57219	1-phosphatidylinos
33	122	6.1	978	2	A70387	conserved hypothet
34	122	6.1	1181	2	C86349	F8K7.4 protein - A
35	121.5	6.1	551	2	S57447	HPBRII-7 protein -
36	121	6.0	792	2	T49989	hypothetical prote
37	120.5	6.0	1902	2	C97702	cell surface antig
38	119.5	6.0	503	1	LUBO11	annexin XI form A
39	119.5	6.0	505	1	S23447	annexin XI form B
40	119	5.9	577	2	T09024	proline-rich prote
41	119	5.9	1357	2	T29265	hypothetical prote
42	119	5.9	1422	2	T24212	hypothetical prote
43	118.5	5.9	384	2	S51796	vasodilator-stimul
44	118.5	5.9	400	2	E70318	hypothetical prote
45	118.5	5.9	915	2	T12526	hypothetical prote

ALIGNMENTS

RESULT 1
H88650
protein C09G12.9 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C;Accession: H88650
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating bio
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www_sanger.ac.uk/Projects/C_
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;
A;Accession: H88650
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-404 <STO>
A;Cross-references: GB:chr_IV; PIDN:AAC25822.1; PID:g3294495; GSPDB:GN00022; CESP:C09
C;Genetics:
A;Gene: C09G12.9
A;Map position: 4

Query Match 32.9%; Score 658; DB 2; Length 404;
Best Local Similarity 36.7%; Pred. No. 5.5e-31;
Matches 151; Conservative 74; Mismatches 121; Indels 66; Gaps 11;

```
Qy          6 KYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNIPICLW 65
              || |   :: :  :: :||| |  |:::| ||  | .  | ||||| |:|  ||||: ::
Db         16 KYADSAKKDIIGALSQFKDLSPGTDTFMFPDGKRRTAFRLKGTIPVYYKGACYNIPVTVY 75

Qy         66 LLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQVMIVVF 125
              | ||:||  |||:| |||:|     :||:  ||::||||:||: |  || ||:|:
Db         76 LWDTHPYYAPICYVNPTSTM---ESEHVNKEGKVFLPYLNEWRFPGYDLSGLLQM----- 127

Qy        126 GDEPPVFSRPISASYPPYQATGP----------PNTSYMPGMPGGISPYPSG-----YP 169
                 :|:|  : |     || |           |  |  | ||   :|||:|     ||
Db        128 -----IFARSAANSATNASATNPSAGSSASSTPTPYPSSQPTMP---TPYPTGSGAAPYP 179

Qy        170 PNPSGYPGC--------------PYPPG----GPYPATTSSQY---PSQPPV---TTVG 204
              |: : ||                |||  |     |||: :|:     |  |||   |:|
Db        180 PSSTPYPSAGAMGYNPYMNVPQSTPYPMGASGASPYPSASSNPAPPPPRPPPVTAQTSVS 239

Qy        205 PSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRAQAELNALKRTEEDLKKGHQKLEEMV 264
              |  |||  ||||||::||| :|:| :::| |     ||: ::: | ::|::| |||: |:
Db        240 SSSGGTIQADTIRASVMSAVEEKIRAKLRERMGTNSAEMASIRTTSDELREGQQKLKRML 299

Qy        265 TRLDQEVAEVDKNIELLKKKDEELSSALEKMENQSENNDIDEVIIPTAPLYKQILNLYAE 324
              |: : : :    |:   |  ||: ||      ::   ||| |    ||::||:  ||:
Db        300 EELETQRSSLQTACEIYTAKKAELAKALSD-AGGTDAPPIDEAIDAAFPLHRQIVLNYAK 358

Qy        325 ENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRKQFQLRALMQKARKTAGL 376
              :           ::|::  | |  :|:||| :||:||  || ||| |:||||
Db        359 DLTC--------QSLKKRQITLAEYLRHVRDVSREQFIYRATMQKCRRTAGL 402
```

RESULT 2
I48283
gene CC2 protein - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 05-Nov-1999
C;Accession: I48283
R;Maucuer, A.; Camonis, J.H.; Sobel, A.
Proc. Natl. Acad. Sci. U.S.A. 92, 3100-3104, 1995
A;Title: Stathmin interaction with a putative kinase and coiled-coil-forming protein dom
A;Reference number: I48282; MUID:95241452; PMID:7724523
A;Accession: I48283
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-83 <RES>
A;Cross-references: EMBL:X82319; NID:g791075; PIDN:CAA57762.1; PID:g791076
C;Genetics:
A;Gene: CC2

Query Match 19.4%; Score 388; DB 2; Length 83;
Best Local Similarity 97.6%; Pred. No. 3.2e-16;
Matches 80; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy      231 RMKEEMDRAQAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSS 290
            ||||||| |||||||||||||||||||||||||||||||||||| |||||||||||||||
Db        1 RMKEEMDGAQAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVGKNIELLKKKDEELSS 60

Qy      291 ALEKMENQSENNDIDEVIIPTA 312
            ||||||||||||||||||||||
Db       61 ALEKMENQSENNDIDEVIIPTA 82
```

RESULT 3
T34520
hypothetical protein DKFZp564J157.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T34520
R;Bloecker, H.; Boecher, M.; Brandt, P.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, October 1999
A;Reference number: Z21539
A;Accession: T34520
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-169 <BLO>
A;Cross-references: EMBL:AL122042
A;Experimental source: fetal brain; clone DKFZp564J157
C;Genetics:
A;Note: DKFZp564J157.1

Query Match 7.7%; Score 153.5; DB 2; Length 169;
Best Local Similarity 30.4%; Pred. No. 0.027;
Matches 42; Conservative 18; Mismatches 45; Indels 33; Gaps 6;

```
Qy      126 GDEPPVFSRPISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNP---SGYPGC---- 178
            |   |    ||:  :||     |      || | |   :||| ||:|    |||||
Db       43 GGSNPAHPPPINPPFPPGPCPPP------PGAPHGNPAFPSGGPPHPVPQPGYPGCQPLG 96

Qy      179 PYPPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDR 238
            ||||  | ||      |  |||  : |   |          : : ||   :|:::| :
Db       97 PYPPPYPPPA------PGIPPVNPLAPGMVGP----------AVIVDK---KMQKKMKK 136

Qy      239 AQAELNALKRTEEDLKKG 256
            |  :::  ::  :  | |
Db      137 AHKKMHKHQKHHKYHKHG 154
```

RESULT 4
LUHU7
annexin VII, long form - human
N;Alternate names: synexin
N;Contains: annexin VII, long form; annexin VII, short form
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1992 #sequence_revision 26-Jan-1996 #text_change 22-Jun-1999
C;Accession: A54467; A32554; A39513; B39513
R;Shirvan, A.; Srivastava, M.; Wang, M.G.; Cultraro, C.; Magendzo, K.; McBride, O.W.;
Biochemistry 33, 6888-6901, 1994
A;Title: Divergent structure of the human synexin (annexin VII) gene and assignment t
A;Reference number: A54467; MUID:94264005; PMID:7515686
A;Accession: A54467
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-488 <SHI>
R;Burns, A.L.; Magendzo, K.; Shirvan, A.; Srivastava, M.; Rojas, E.; Alijani, M.R.; P
Proc. Natl. Acad. Sci. U.S.A. 86, 3798-3802, 1989
A;Title: Calcium channel activity of purified human synexin and structure of the huma
A;Reference number: A32554; MUID:89264510; PMID:2542947
A;Accession: A32554
A;Molecule type: mRNA
A;Residues: 1-145,168-488 <BUR>
A;Cross-references: EMBL:J04543; NID:g338243; PIDN:AAA36616.1; PID:g338244
R;Magendzo, K.; Shirvan, A.; Cultraro, C.; Srivastava, M.; Pollard, H.B.; Burns, A.L.
J. Biol. Chem. 266, 3228-3232, 1991
A;Title: Alternative splicing of human synexin mRNA in brain, cardiac, and skeletal m
A;Reference number: A39513; MUID:91131630; PMID:1825209
A;Accession: A39513
A;Molecule type: mRNA
A;Residues: 137-145,168-176 <MAG>
A;Cross-references: EMBL:J05732
A;Accession: B39513
A;Molecule type: mRNA
A;Residues: 137-176 <MA2>
C;Comment: Annexins undergo reversible, calcium-dependent binding to membrane phospho
derstood.
C;Comment: The long form of annexin VII is more prevalent in brain, heart, and skelet
C;Genetics:
A;Gene: GDB:ANX7
A;Cross-references: GDB:369042; OMIM:186360
A;Map position: 10q21.1-10q21.2
A;Introns: 17/3; 87/1; 124/1; 145/3; 167/3; 202/2; 233/3; 271/3; 328/3; 385/3; 411/1;
C;Superfamily: annexin VII; annexin repeat homology
C;Keywords: alternative splicing; calcium binding; calcium channel; duplication; endo
F;1-488/Product: annexin VII, long form #status predicted <MAT>
F;1-145,168-488/Product: annexin VII, short form #status predicted <MA3>
F;188-259/Domain: annexin repeat homology <AX1>
F;199-215/Region: endonexin fold #status predicted
F;260-331/Domain: annexin repeat homology <AX2>
F;271-287/Region: endonexin fold #status predicted
F;343-415/Domain: annexin repeat homology <AX3>
F;355-371/Region: endonexin fold #status predicted
F;419-488/Domain: annexin repeat homology <AX4>
F;430-446/Region: endonexin fold #status predicted

Query Match 7.3%; Score 147; DB 1; Length 488;
Best Local Similarity 21.7%; Pred. No. 0.23;
Matches 89; Conservative 34; Mismatches 88; Indels 200; Gaps 19;

```
Qy      139 SYPPYQATG-PPNTSYMPG------MPGGISPYPSGYP---------------------- 169
            ||| |  || ||   | |        | |  |||||:|
Db        2 SYPGYPPTGYPPFPGYPPAGQESSFPPSGQYPYPSGFPPMGGGAYPQVPSSGYPGAGGYP 61

Qy      170 -----PNPSGYPGCPYPPGGP-YPATTSSQ-------------YPSQPPVTTVG------ 204
                | | |||| | | | | ||      |              || |||  : |
Db       62 APGGYPAPGGYPGAPQPGGAPSYPGVPPGQGFGVPPGGAGFSGYP-QPPSQSYGGGPAQV 120

Qy      205 ----------------------PSRDGT-------------------------ISEDTIR 217
                                   ||:  |                         ::: |||
Db      121 PLPGGFPGGQMPSQYPGGQPTYPSQINTDSFSSYPVFSPVSLDYSSEPATVTQVTQGTIR 180

Qy      218 -ASLISAVSDK--LRWRMK--------------EEMDRAQAELNALKRTE-----EDLK 254
             |:   |: |   ||  ||                  |: |    | | :       :|||
Db      181 PAANFDAIRDAEILRKAMKGFGTDEQAIVDVVANRSNDQRQKIKAAFKTSYGKDLIKDLK 240

Qy      255 K------------------------------GHQK---LEEMVTRLDQEVAEV----- 274
                                           | |:   :| : || :||: |:
```

```
Db        241 SELSGNMEELILALFMPPTYYDAWSLRKAMQGAGTQERVLIEILCTRTNQEIREIVRCYQ 300

Qy        275 -DKNIELLKKKDEELSSALEKM----------ENQSENNDIDEVIIPTAPLYKQILNLYA 323
              :  :| |    : |   |::          |||| |: :                   |
Db        301 SEFGRDLEKDIRSDTSGHFERLLVSMCQGNRDENQSINHQM-----------------A 342

Qy        324 EENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRKQFQLRALMQKARKTA 374
              :|:|    ::  ||    | :  |    :: | :|   |||| |:   : |
Db        343 QEDA--QRLYQAGE----GRLGTDESCFNMILATRSFPQLRATMEAYSRMA 387


RESULT 5
S74288
hypothetical protein YCL008c - yeast (Saccharomyces cerevisiae)
C;Species: Saccharomyces cerevisiae
C;Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 19-Apr-2002
C;Accession: S74288; S19410
R;Wedler, H.; Wambutt, R.
submitted to the Protein Sequence Database, September 1996
A;Reference number: S74288
A;Accession: S74288
A;Molecule type: DNA
A;Residues: 1-212 <WED>
A;Cross-references: EMBL:X59720; NID:g1907116; PIDN:CAA42351.1; PID:e309039; PID:g190713
R;Oliver, S.G.; Anwar, R.; Brown, A.; Gent, M.E.; Indge, K.J.; James, C.M.; Stateva, L.I
submitted to the Protein Sequence Database, March 1992
A;Reference number: S19337
A;Accession: S19410
A;Molecule type: DNA
A;Residues: 1-90,'GEYS',95,'TA',98-99,'AQVSTP',106,'TAIASTSTTSTGI' <OLI>
A;Cross-references: EMBL:X59720; MIPS:YCL008c
A;Note: this sequence has been revised in reference S74288
C;Genetics:
A;Gene: SGD:STP22
A;Cross-references: SGD:S0000514
A;Map position: 3L
A;Note: YCL008c

  Query Match             7.3%;  Score 145.5;  DB 2;  Length 212;
  Best Local Similarity   24.5%;  Pred. No. 0.1;
  Matches   58;  Conservative   38;  Mismatches  100;  Indels   41;  Gaps   11;

Qy         64 LWLLDTYPYNPPIC------FVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGL 117
              :|:   ||  ||         |  | | ::   :::|:|| | || || |     :|: :
Db          1 MWVPSMYPVKPPFISINLENFDMNTISSSLPIQEYIDSNGWIALPILHCWDPAAMNLIMV 60

Qy        118 IQVMIVVFGDEPPVFSRPISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPG 177
              :| :: :   |||    |     |  |     ||  :|  |      |   || |
Db         61 VQELMSLL-HEPPQDQAPSLPPKPNTQLQQEQNTPPLPPKPKS----PHLKPPLP----- 110

Qy        178 CPYPPGGPYPATTSSQYPSQPPVTTVGPSRD-------GTISEDTIRASLISAVSDKLRW 230
                ||  | ||: :         | : |:          |:  :  |   :  |:||:
Db        111 ---PPPPPQPASNALDLMDMDN-TDISPTNHHEMLQNLQTVVNELYRED-VDYVADKILT 165

Qy        231 R---MKEEMDRAQAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKK 284
              |   |:| : |   |: |       : | | :  |    ::| :  ::  |::| :|
Db        166 RQTVMQESIARFH-EIIA-------IDKNHLRAVEQA--IEQTMHSLNAQIDVLNRK 212


RESULT 6
A53152
annexin XI - human
C;Species: Homo sapiens (man)
C;Date: 07-Jul-1995 #sequence_revision 07-Jul-1995 #text_change 13-Aug-1999
C;Accession: A53152
R;Misaki, Y.; Pruijn, G.J.M.; van der Kemp, A.W.C.M.; van Venrooij, W.J.
J. Biol. Chem. 269, 4240-4246, 1994
A;Title: The 56K autoantigen is identical to human annexin XI.
A;Reference number: A53152; MUID:94140847; PMID:7508441
A;Accession: A53152
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-505 <MIS>
A;Cross-references: GB:L19605; NID:g457128; PIDN:AAA19734.1; PID:g457129
C;Genetics:
A;Gene: GDB:ANX11
A;Cross-references: GDB:313076
A;Map position: 9q11-9q22
C;Superfamily: annexin VII; annexin repeat homology
C;Keywords: calcium binding; duplication; endonexin fold; glycoprotein; phospholipid
F;203-274/Domain: annexin repeat homology <AX1>
F;275-346/Domain: annexin repeat homology <AX2>
F;358-430/Domain: annexin repeat homology <AX3>
F;434-505/Domain: annexin repeat homology <AX4>

  Query Match             7.1%;  Score 142;  DB 2;  Length 505;
  Best Local Similarity   28.0%;  Pred. No. 0.48;
  Matches   63;  Conservative   19;  Mismatches   91;  Indels   52;  Gaps   10;

Qy        111 QSDLLGLIQVMIVVFG--DEPPVFSRPISASYPPYQ--ATGPPNTSYMPGMPGGISPYPS 166
              |  | |:   |   ||  : | ::     | |||      | | ::  |  | |: | |
Db         51 QDYLSGMAANMSGTFGGANMPNLYPGAPGAGYPPVPPGGFGQPPSAQQPVPPYGMYPPPG 110

Qy        167 GYPPN--PS--GYPGCPYP----------PGGPYPATTSSQYPSQPPVTTVG-----PSR 207
              | ||:  ||   ||| | |          | | ||      || ||||   |     ||
Db        111 GNPPSRMPSYPPYPGAPVPGQPMPPPGQQPPGAYPGQPPVTYPGQPPVPLPGQQQPVPSY 170

Qy        208 DGTISEDTI----------RASLISAV-------SDKLRWRMKEEMDRAQAELNALKRT 249
               |     |:          | ::  |        :: ||  ||      || :: |
Db        171 PGYPGSGTVTPAVPPTQFGSRGTITDAPGFDPLRDAEVLRKAMKGFGTDEQAIIDCLGSR 230

Qy        250 EEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEK 294
                |:  | |    |   :         :|:|    |||   ||
Db        231 SN--KQRQQILLSFKTAYGK---------DLIKDLKSELSGNFEK 264


RESULT 7
S29170
annexin VII - mouse
N;Alternate names: synexin
C;Species: Mus musculus (house mouse)
C;Date: 25-Feb-1994 #sequence_revision 01-Sep-1995 #text_change 13-Aug-1999
C;Accession: S29170; S46209; S51173
R;Zhang-Keck, Z.Y.; Burns, A.L.; Pollard, H.B.
Biochem. J. 289, 735-741, 1993
A;Title: Mouse synexin (annexin VII) polymorphisms and a phylogenetic comparison with
A;Reference number: S29170; MUID:93168121; PMID:7916616
A;Accession: S29170
A;Molecule type: mRNA
A;Residues: 1-463 <ZHA>
A;Cross-references: EMBL:L13129
R;Zhang-Keck, Z.Y.; Srivastava, M.; Kozak, C.A.; Caohuy, H.; Shirvan, A.; Burns, A.L.
Biochem. J. 301, 835-845, 1994
A;Title: Genomic organization and chromosomal localization of the mouse synexin gene.
A;Reference number: S46209; MUID:94330961; PMID:8053909
A;Accession: S46209
A;Molecule type: DNA
A;Residues: 1-463 <ZHF>
R;Zhang-Keck, Z.Y.; Burns, A.L.; Pollard, H.B.
submitted to the EMBL Data Library, May 1993
A;Reference number: S51173
A;Accession: S51173
A;Molecule type: mRNA
A;Residues: 1-144,'S',146-303,'A',305-463 <ZHW>
A;Cross-references: EMBL:L13129; NID:g293293; PIDN:AAA37238.1; PID:g293294
C;Genetics:
A;Gene: MGI:Anx7
A;Cross-references: MGI:88031
A;Map position: 14
A;Introns: 19/3; 87/1; 121/1; 142/3; 177/1; 208/3; 246/3; 303/3; 360/3; 386/1; 423/3
C;Superfamily: annexin VII; annexin repeat homology
C;Keywords: calcium channel; ion channel
F;163-234/Domain: annexin repeat homology <AX1>
F;235-306/Domain: annexin repeat homology <AX2>
F;318-390/Domain: annexin repeat homology <AX3>
F;394-463/Domain: annexin repeat homology <AX4>
```

```
  Query Match             7.0%;  Score 141;  DB 2;  Length 463;
  Best Local Similarity   22.0%;  Pred. No. 0.49;
  Matches   76;  Conservative   43;  Mismatches   90;  Indels  136;  Gaps   17;

Qy        139 SYPPYQATGPPNTSYMP---------GMPGGISP-----YPSGY----PPNPSGYPGCPY 180
              :|||  : | |      |          | || :||     || |     ||  :|: | |
Db         45 AYPPAPSGGYPGAGGYPAPGGYPAPGGYPGALSPGGPPAYPGGQGFGAPPGGAGFSGYPQ 104

Qy        181 PP-----GGP-----------------YPATTSSQYPSQPPVTTVGPSRDGTI----SED 214
              ||     |||                ||    : |||||   | |    |||    : |
Db        105 PPAQSYGGGPAQVPVPGGFPGGQMPSQYPG-GQAPYPSQPAAMTQG--TQGTILPASNFD 161

Qy        215 TIRASLI-----------SAVSDKLRWRMKEEMDRAQAELNAL--KRTEEDLKK----- 255
               :| : |            |: | :  |  ::  : :|    :  |   :|||
Db        162 AMRDAEILRKAMKGFGTDEQAIVDVVSNRSNDQRQQIKAAFKTMYGKDLIKDLKSELSGN 221

Qy        256 --------------------------GHQK---LEEMVTRLDQEVAEVDKNIELLKKKD 285
                                         | |:   :| : || :||: :: :  :|   :|
Db        222 MEELILALFMPSTYYDAWSLRKAMQGAGTQERVLIEILCTRTNQEIRDIVRCYQLEFGRD 281

Qy        286 ------EELSSALEKM----------ENQSENNDIDEVIIPTAPLYKQILNLYAEENAIE 329
                   : |   |::         | || |: :                 |:|:|
Db        282 LEKDIRSDTSGHFERLLVSMCQGNRDERQSVNHQM-----------------AQEDA-- 321

Qy        330 DTIFYLGEALRRGVIDLDVFLKHVRLLSRKQFQLRALMQKARKTA 374
                ::  ||    | :  |    :: | :|   ||:| |:   : |
Db        322 QRLYQAGE----GRLGTDESCFNMILATRSFPQLKATMEAYSRMA 362


RESULT 8
T39608
zinc finger transcription factor - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 02-Sep-2000
C;Accession: T39608
R;Lyne, M.; Wood, V.; Rajandream, M.A.; Barrell, B.G.; Beck, A.; Reinhardt, R.
submitted to the EMBL Data Library, May 1998
A;Reference number: Z21866
A;Accession: T39608
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-827 <LYN>
A;Cross-references: EMBL:AL023554; PIDN:CAA19035.1; GSPDB:GN00067; SPDB:SPBC16G5.16
A;Experimental source: strain 972h-; cosmid c16G5
C;Genetics:
A;Gene: SPDB:SPBC16G5.16
A;Map position: 2
A;Introns: 30/3; 41/1
C;Superfamily: GAL4 zinc binuclear cluster homology
F;11-47/Domain: GAL4 zinc binuclear cluster homology <GL4>

  Query Match             6.9%;  Score 138.5;  DB 2;  Length 827;
  Best Local Similarity   25.2%;  Pred. No. 1.4;
  Matches   60;  Conservative   28;  Mismatches   99;  Indels   51;  Gaps   11;

Qy         13 RETVNVITLYKDLKPVLDSYVFNDGSSREL---MNLTGTIPVPYRGNTYNIP-----ICL 64
              ||:    |: : :| :| :|  |    :  |    ::|| :|  :    | |||       :
Db        589 RESTYYSTIMEAIKNLLIAYDMNSSGTENLDATPDVTGQLPNNFSQRTSNIPREFPQAQI 648

Qy         65 WLLDT-YP--YNP------PICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLL 115
              :  |  ||  |||      |  |  ||     :   :   ||  |  |     :|   ::
Db        649 FYSDAPYPGYYNPAQFQNAPTNFPMPTYGGRTQDQSYPRQNG--YPSYSDGNVYPHDRVM 706

Qy        116 GLIQVMIVVFGDEPPVFSR------------PISASYPPYQATGPPNTSYMPGMPGGISP 163
                  : :|   |  :              | : ||||| :     || ||      |
Db        707 -------INYGSSMPTANGFYVPNTYSPVPFPYNTSYPPYMSP----TSNMPQAFQAYSQ 755

Qy        164 YPSGYPPNPSGYPGCPYPPGG----PYPATTSSQYPSQPPVTTVGPSRDGTISEDTIR 217
              ||  :|| |      | |  |    |  | :   ||    |      ||    ::  |:|
Db        756 YPYQHPPFPLSEQMLPLPTSGVMMAPGAAKSGMPYPFIQP-----PSMTNQVAYPTVR 808


RESULT 9
LURB11
annexin XI - rabbit
N;Alternate names: calcyclin-associated annexin protein CAP-50
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Jun-2000
C;Accession: JH0694; PH0950; A38250; PS0263
R;Tokumitsu, H.; Mizutani, A.; Muramatsu, M.; Yokota, T.; Arai, K.; Hidaka, H.
Biochem. Biophys. Res. Commun. 186, 1227-1235, 1992
A;Title: Molecular cloning of rabbit CAP-50, a calcyclin-associated annexin protein.
A;Reference number: JH0694; MUID:92378579; PMID:1380798
A;Accession: JH0694
A;Molecule type: mRNA
A;Residues: 1-503 <TOK>
A;Cross-references: DDBJ:D10883; NID:g471147; PIDN:BAA01705.1; PID:g471148
A;Experimental source: lung
A;Accession: PH0950
A;Molecule type: protein
A;Residues: 104-141;213-231;254-262;270-280;285-309;319-337;429-448;478-492 <TK2>
R;Tokumitsu, H.; Mizutani, A.; Minami, H.; Kobayashi, R.; Hidaka, H.
J. Biol. Chem. 267, 8919-8924, 1992
A;Title: A calcyclin-associated protein is a newly identified member of the Ca(2+)/ph
A;Reference number: A38250; MUID:92250478; PMID:1533622
A;Accession: A38250
A;Molecule type: protein
A;Residues: 104-141;213-223,'X',225-231;254-263;271-280;285-291,'X',293-300,'X',302-3
C;Comment: This protein binds specifically to calcyclin in a Ca2+ dependent manner.
C;Superfamily: annexin VII; annexin repeat homology
C;Keywords: calcium binding; duplication; endonexin fold; glycoprotein; phospholipid
F;201-272/Domain: annexin repeat homology <AX1>
F;212-228/Region: endonexin fold #status predicted
F;273-344/Domain: annexin repeat homology <AX2>
F;284-300/Region: endonexin fold #status predicted
F;356-428/Domain: annexin repeat homology <AX3>
F;368-384/Region: endonexin fold #status predicted
F;432-503/Domain: annexin repeat homology <AX4>
F;443-459/Region: endonexin fold #status predicted
F;58/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             6.8%;  Score 136.5;  DB 1;  Length 503;
  Best Local Similarity   24.1%;  Pred. No. 0.98;
  Matches   81;  Conservative   39;  Mismatches  117;  Indels   99;  Gaps   18;

Qy        126 GDEPPV----FSRPISASYPPYQATGPPNTSYMPGMPGGISPYPSG---YPPNPSG-YPG 177
              |  |||    | :|     || | : ||   | |  ||| :| |||   ||| |    ||
Db         79 GGYPPVPPGGFGQP-----PPTQPSVPPYGVYPP--PGG-NP-PSGVPSYPPFPGAPVPG 129

Qy        178 CPYP-----PGGPYPATTSSQYPSQPPVTTVG----PSRDGTISEDTI-----------R 217
              | |      | ||||      || | ||   |    ||  |     |:           |
Db        130 QPMPPPGHQPPGPYPGQLPVTYPGQSPVPPPGQQPMPSYPGYPGSGTVTPAVPPVQFGNR 189

Qy        218 ASLISAV-------SDKLRWRMKEEMDRAQAELNAL-----KRTEEDLKK-----GHQKL 260
              ::  |        :: ||  ||       || :: |     |: :: |       |   :
Db        190 GTITDASGFDPLRDAEVLRKAMKGFGTDEQAIIDCLGSRSNKQRQQILLSFKTAYGKDLI 249

Qy        261 EEMVTRLD------------------QEVAEVDKN--------IELLKKKDEELSSALE 293
              ::: : |                  |: |  |          ||:|  :  |    |
Db        250 KDLKSELSGNFEKTILALMKTPILFDAYEIKEAIKGAGTDEACLIEILASRSNEHIRELN 309

Qy        294 KMENQSENNDIDEVI-IPTAPLYKQILNLYAEENAIEDT-------------IFYLGEAL 339
              |       ::| |   |:  ::::|   :: |  | |              ::  ||
Db        310 KAYKTEFKKTLEEAIRSDTSGHFQRLLISLSQGNRDESTNVDMSLVQRDVQELYAAGE-- 367

Qy        340 RRGVIDLDVFLKHVRLLSRKQFQLRALMQKARKTAG 375
                 :  |    :  | || :  | |:  : ::  |
Db        368 --NRLGTDESKFNAVLCSRSRAHLVAVFNEYQRMTG 401


RESULT 10
T14192
extensin homolog T28D5.70 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-Jan-2000
```

```
C;Accession: T14192
R;Bevan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Barrell, B.G.; Bancrc
submitted to the Protein Sequence Database, August 1999
A;Reference number: Z17931
A;Accession: T14192
A;Molecule type: DNA
A;Residues: 1-437 <BEV>
A;Cross-references: EMBL:AL109819
A;Experimental source: cultivar Columbia; BAC clone T28D5
C;Genetics:
A;Gene: ATSP:T28D5.70
A;Map position: 4
C;Superfamily: hydroxyproline-rich glycoprotein

  Query Match             6.8%;  Score 135.5;  DB 2;  Length 437;
  Best Local Similarity   26.8%;  Pred. No. 0.95;
  Matches   49;  Conservative   16;  Mismatches   75;  Indels   43;  Gaps    8;

Qy         50 PVPYRGNTYNIPICLWLLDTYPY----NPPICFVKPTSSMTIKTGKHVDANGKIYL---- 101
              | ||   ||: |   ::  : ||     ||  :  | |    |:  :| ::   |
Db         43 PPPY---TYSPPPSPYVYKSPPYVYSSPPPYAYSPPPSPYVYKSPPYVYSSPPPYAYSPP 99

Qy        102 --PYLHEWKHPQSDLLGLIQVMIVVFGDEPPVFSRPISASY----PPYQATGPPNTSYMP 155
                ||:  :| |              |:   ||    |  : |    |||  : ||   |
Db        100 PSPYV--YKSPP-----------YVYSSPPPYAYSPPPSPYVYKSPPYVYSSPPPYVYSS 146

Qy        156 GMPGGISPYPSGY--PPNPSGYPGCPY-----------PPGGPYPATTSSQYPSQPPVTT 202
                |   || |  |  ||:|  |   ||           ||  ||    :     | ||
Db        147 PPPYAYSPPPYAYSPPPSPYVYKSPPYVYSSPPPYAYSPPPSPYVYKSPPYVYSSPPPYA 206

Qy        203 VGP 205
                |
Db        207 YSP 209


RESULT 11
D70509
hypothetical protein Rv1233c - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C;Accession: D70509
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: D70509
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-198 <COL>
A;Cross-references: GB:Z98260; GB:AL123456; NID:g3261826; PIDN:CAB10930.1; PID:e1299942;
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv1233c

  Query Match             6.7%;  Score 134;  DB 2;  Length 198;
  Best Local Similarity   32.8%;  Pred. No. 0.44;
  Matches   42;  Conservative   12;  Mismatches   42;  Indels   32;  Gaps    7;

Qy        126 GDEPPVFSRPIS---------------ASYPP--YQATGPPNTSYMPGMPGGISP---- 163
                |  |||  ||               |::||  |  :| |  :| || | |  |
Db         17 GGPPPVGERPPEQPIADAPWAPPASSPMANHPPPAYPPSGYP-PAYQPGYPTGYPPPMPP 75

Qy        164 ---YPSGYPP---NPSGYPGCPYPP-GGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTI 216
                 | ||||   : :||   ||||   ||  :    ||    :  |||: |  :   :
Db         76 GGYAPPGYPPPGTSSAGYGDIPYPPMPPPYGGSPGGYYPEPGYLDGYGPSQPGMNTMALV 135

Qy        217 RASLISAV 224
                 |||||:
Db        136 --SLISAL 141


RESULT 12
T37550
hypothetical coiled-coil protein [imported] - fission yeast (Schizosaccharomyces pomb
C;Species: Schizosaccharomyces pombe
C;Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 08-Sep-2000
C;Accession: T37550
R;Connor, R.; Churcher, C.M.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, October 1999
A;Reference number: Z21723
A;Accession: T37550
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-485 <CON>
A;Cross-references: EMBL:AL132714; PIDN:CAB59797.1; GSPDB:GN00066; SPDB:SPAC11H11.01
A;Experimental source: strain 972h-; cosmid c11H11
C;Genetics:
A;Gene: SPDB:SPAC11H11.01
A;Map position: 1

  Query Match             6.7%;  Score 134;  DB 2;  Length 485;
  Best Local Similarity   19.8%;  Pred. No. 1.3;
  Matches   83;  Conservative   61;  Mismatches  139;  Indels  136;  Gaps   15;

Qy         63 CLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQS--DLLGLIQV 120
              |: :   ||  |||          :  | |||     |     |  ||: :|  ::  |:|
Db         79 CITVRSYYPSQPPI-----VQLLDEKGGKH-----KYTSLLLQLWKNSRSVFNIYRLVQA 128

Qy        121 MI--------------------------VVFGDEPPVFSRPISASYPPYQATGPPNTSY 153
              :|                                : |||          |: ::   :
Db        129 LIKQDFEREHTSPPELPTKLVNTIEKLKVKEENEAPPVIPA------KPFSSSSEQHFRK 182

Qy        154 MPGMPGGISPYP----------SGYPPNPSGYPGCPYPPG-GPYPATT------------ 190
              :| :|  : | |                | | :    :     |: ||:
Db        183 VPALPSKLPPKPLKITANSSLGQETNSNSSSFQSTLFSLNTAPFSATSQQLVHDSVSLRR 242

Qy        191 -SSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDR----------- 238
               ||  |:| |:        :  |::||   ||    |       | :: :
Db        243 PSSNIPAQKPIPPKPEQNEIIITKDT--PSLKDKYSKPALLPQKPKVSKGQIVQQVSVFS 300

Qy        239 ------AQAELNALKRTEEDLKKGHQKL--------------------------EEMVT 265
                   :|: || :    |   ||  :|                          |: ::
Db        301 TGKKIESQSLLNLIDTDIETPLKGSSELLYSEDFKPNVDPVKIQQILHKQNKIIEEKWIS 360

Qy        266 R-------------LDQE---VAEVDKNIE----LLKKKDEELSSALEKMENQSENNDID 305
              :              ||||   :  : ||||    :| |:  :   ||:|::|  : :  :
Db        361 QIRISKNLEVKQRLLDQERHALETLAKNIENNRFILGKRRRKAREALQKLDNLKDLSVQE 420

Qy        306 EVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLLDVFLKHVRLLSRKQFQLR 364
               |||:    |    |  ::  ::: |  | :||    |    :|| ::||:|:|| :|
Db        421 LFIIPSERELK-YYELKRKDEKLDEGIRALNQALHHESIMPASWLKGIKLLARQQFLIR 478


RESULT 13
T36037
probable export associated protein - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T36037
R;Seeger, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A
submitted to the EMBL Data Library, March 1999
A;Reference number: Z21581
A;Accession: T36037
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-671 <SEE>
A;Cross-references: EMBL:AL035591; PIDN:CAB38154.1; GSPDB:GN00070; SCOEDB:SCC54.24c
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SCC54.24c

  Query Match             6.6%;  Score 132.5;  DB 2;  Length 671;
  Best Local Similarity   27.8%;  Pred. No. 2.4;
```

```
  Matches   59;  Conservative   25;  Mismatches   73;  Indels   55;  Gaps   14;

Qy        126 GDEPPVFSRPISASYPPYQATGP----PNTSYMPGMPGGISPYPSGYPPNP--SGYPGCP 179
              |  ||           ||   ||    | |   || |||  |||||| |  |  :| |  |
Db        212 GPVPPA---------SPYDPQGPSAPGPQTPGAPGRPGGQQPYPSG-PGEPATAGQPSAP 261

Qy        180 YPPGGP----YPATTSSQYPSQPPVTTVG----PSR----DGT--------------ISE 213
                |||    :|   |:  |  ||          |:|    | |              : |
Db        262 -AAGGPIDYSHPRRQSAAPPPPPPAAPASEPGRPARPVAGDATGWSMEERLYNQVWGMFE 320

Qy        214 DTIR--ASLISAVSDKLRWRMKEEMDRAQAELNAL-----KRTEEDLKKGHQKLEEMVTR 266
              |  |  |:  |||  |    ||::|:|:  ::  :          |  :  | :|
Db        321 DLARTTAAYRSAV-DFADSRMEKELDQVLSDPRSRIGGQGDAAREAARARHGQLVSQARE 379

Qy        267 -LDQEVAEVDKNIELLKKKDEELSSALEKMEN 297
               ||::||::    |::   :  | :|  : :|
Db        380 VLDRDVAQLVAEAEVV---EPALPAAFARWDN 408


RESULT 14
T42532
hypothetical protein - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000
C;Accession: T42532
R;Yoshioka, S.; Kato, K.; Nakai, K.; Okayama, H.; Nojima, H.
DNA Res. 4, 363-369, 1997
A;Title: Identification of open reading frames in Schizosaccharomyces pombe cDNAs.
A;Reference number: Z17323; MUID:98162722; PMID:9501991
A;Accession: T42532
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-370 <YOS>
A;Cross-references: EMBL:D89169; NID:g1749545; PIDN:BAA13831.1; PID:g1749546
A;Experimental source: strain PR745

  Query Match             6.6%;  Score 131.5;  DB 2;  Length 370;
  Best Local Similarity   22.0%;  Pred. No. 1.3;
  Matches   76;  Conservative   42;  Mismatches  145;  Indels   83;  Gaps   15;

Qy          8 RDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTI--PVPYRGNTYNIPICLW 65
              :|| : |     :    ::|  |  :    | :   |   |   | | : | | |
Db         72 KDLRIEEPATTPSA-PPVQPPNDPAIIGSNSGQYNWNQAQTAQPPQPVQPNPYGAPYQQA 130

Qy         66 LLDTYPY----NPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQVM 121
                   ||    | |  || |     :  |  :||:     | :  :  |      :|
Db        131 PPAGAPYYMYPNAPAQFVPPGG---LPLGTPLDAS----TPAVPYYGAPDQQQMG----- 178

Qy        122 IVVFGDEPPVFSRPISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCPYP 181
                : | |:: :|    :   | |    |||   |        |:  || | |
Db        179 ------QRPEFAQNVSQG---FAGQAPYNVRPGYGMPSNQKP--------PNFAPGMPAP 221

Qy        182 PGGPYPATTSSQYPSQPPVTTVGPS---------------RDGTISED-----TIRASL 220
              ||   : |    | ||   | |                 |: |:: |     |:  |
Db        222 --GPTAVSASPSLQSMPPTNGVIPGAQPSIEASIEKESTSIRNSTVTNDRVVNTTVDVSQ 279

Qy        221 ISAVSDKLRWRMKEEMDRAQAELNALK-RTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIE 279
                |         :|:   | : :| | ||| | :  :|| : |              :
Db        280 SQTVETS---GPSKEVPTTQPDASAAKPRTELDFQTANQKFQSM------------KDD 323

Qy        280 LLKKKDEELSSALEKMENQSENNDIDEVIIPTAPLYKQILNLYAEE 325
              ||| |::| :    | :          :||    : :::  |  ||
Db        324 LLKGKNDEEAEEFHKPKQSF-------LIILVVKVKRRVWKLQIEE 362


RESULT 15
S39059
protein BRG1 - human
C;Species: Homo sapiens (man)
C;Date: 25-Feb-1994 #sequence_revision 17-Nov-1995 #text_change 02-Aug-2002
C;Accession: S39059
R;Khavari, P.A.; Peterson, C.L.; Tamkun, J.W.; Mendel, D.B.; Crabtree, G.R.
Nature 366, 170-174, 1993
A;Title: BRG1 contains a conserved domain of the SWI2/SNF2 family necessary for norma
A;Reference number: S39059; MUID:94050144; PMID:8232556
A;Accession: S39059
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1613 <KHA>
C;Superfamily: human SNF2alpha protein; bromodomain homology
F;1451-1506/Domain: bromodomain homology <BRO>

  Query Match             6.5%;  Score 131;  DB 2;  Length 1613;
  Best Local Similarity   19.7%;  Pred. No. 8.4;
  Matches   73;  Conservative   44;  Mismatches  115;  Indels  138;  Gaps   13;

Qy        126 GDEPPVFSRPISASYPPYQATGPPNTSYMPGMP-----GGISPYPSG------------- 167
              |  || :|||     |     ||  :   ||||     |   |:| |
Db        242 GPAPPNYSRPHGMGGPNMPPPGP--SGVPPGMPGQPPGGPPKPWPEGPMANAAAPTSTPQ 299

Qy        168 --YPPNPSGYPGCPYPPGGP-------YPATTSSQYPSQP--------------------- 198
                  || |:| |  | ||  |        | | |   |:||
Db        300 KLIPPQPTGRPS-PAPPAVPPAASPVMPPQTQSPGQPAQPAPMVPLHQKQSRITPIQKPR 358

Qy        199 ----------------------------PVTTVGPSRDGTISE---------------- 213
                                         | : ||  |     |
Db        359 GLDPVEILQEREYRLQARIAHRIQELENLPGSLVGDLRTKVTIELKDLRLLNFQRQLRQE 418

Qy        214 -------DTI--------------RASLISA-VSDKLRWRMKEEMDRA-----QAELNAL 246
                    ||               | ||  | :::||  : | | :|      |  ||::
Db        419 VVVCMRRDTALETALNVKAYKRSKRQSLREARITEKLEKQQKIEQERKRRQKHQEYLNSI 478

Qy        247 KRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQSENNDIDE 306
               :  :| |: |: :   : :| : |:    | |  :||: |          : |   |
Db        479 LQHAKDFKEYHRSVTGTIQKLTKAVSTYHANTEREQKKENE----------RIEKERIRR 528

Qy        307 VIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDL-DVFLKHVRLLSRKQFQLRA 365
              ::      |:::::   ::      : || :    | :| ::  :|      |: : :
Db        529 LMAEDEEGYRKLIDQKKDKR-----LAYLLQQTDEYVANLTELVRQHKASQVAKEIKKKK 583

Qy        366 LMQKARKTAG 375
               :||    |
Db        584 TKKKAENAEG 593


Search completed: June  3, 2003, 14:09:58
Job time : 27.9658 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: June 3, 2003, 14:05:55 ; Search time 13.9816 Seconds
(without alignments)
1127.267 Million cell updates/sec

Title: US-09-804-690-4
Perfect score: 2002
Sequence: 1 MVSKYKYRDLTVRETVNVIT.....FQLRALMQKARKTAGLSDLY 380

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2002	100.0	390	1	T101_HUMAN	Q99816 homo sapien
2	1900.5	94.9	391	1	T101_MOUSE	Q61187 mus musculu
3	160	8.0	466	1	ANX7_HUMAN	P20073 homo sapien
4	145.5	7.3	212	1	YCA8_YEAST	P25604 saccharomyc
5	142	7.1	505	1	ANXB_HUMAN	P50995 homo sapien
6	141.5	7.1	503	1	ANXB_MOUSE	P97384 mus musculu
7	141	7.0	463	1	ANX7_MOUSE	Q07076 mus musculu
8	136.5	6.8	503	1	ANXB_RABIT	P33477 oryctolagus
9	130.5	6.5	338	1	FOSB_HUMAN	P53539 homo sapien
10	130	6.5	1647	1	SN24_HUMAN	P51532 homo sapien
11	127	6.3	1185	1	DRPL_HUMAN	P54259 homo sapien
12	126	6.3	1453	1	CA11_MOUSE	P11087 mus musculu
13	124	6.2	964	1	YQY1_CAEEL	Q09560 caenorhabdi
14	123.5	6.2	380	1	VASP_HUMAN	P50552 homo sapien
15	123	6.1	1248	1	DIA1_HUMAN	O60610 homo sapien
16	122.5	6.1	746	1	PCAP_HUMAN	Q96rn5 homo sapien
17	122	6.1	978	1	RA50_AQUAE	O67124 aquifex aeo
18	120.5	6.0	376	1	VASP_MOUSE	P70460 mus musculu
19	120.5	6.0	753	1	ZIN_HUMAN	Q9nrl3 homo sapien
20	120	6.0	760	1	ZIN_MOUSE	P58404 mus musculu
21	120	6.0	1101	1	DIA2_HUMAN	O60879 homo sapien
22	119.5	6.0	503	1	ANXB_BOVIN	P27214 bos taurus
23	119.5	6.0	1224	1	DYNA_CHICK	P35458 gallus gall
24	119	5.9	1183	1	DRPL_RAT	P54258 rattus norv
25	119	5.9	1280	1	DYNA_RAT	P28023 rattus norv
26	118.5	5.9	384	1	VASP_CANFA	P50551 canis famil
27	118.5	5.9	792	1	PCAP_MOUSE	Q924h2 mus musculu
28	118.5	5.9	1255	1	DIA1_MOUSE	O08808 mus musculu
29	118.5	5.9	1281	1	DYNA_MOUSE	O08788 mus musculu
30	118	5.9	1171	1	DIA3_MOUSE	Q9z207 mus musculu
31	118	5.9	1460	1	CA11_CANFA	Q9xsj7 canis famil
32	117.5	5.9	620	1	EXTN_TOBAC	P13983 nicotiana t
33	117	5.8	505	1	WASL_BOVIN	Q95107 bos taurus
34	116.5	5.8	338	1	FOSB_MOUSE	P13346 mus musculu
35	116.5	5.8	859	1	MUTS_AQUAE	O66652 aquifex aeo
36	116.5	5.8	1790	1	USO1_YEAST	P25386 saccharomyc
37	116	5.8	462	1	ANX7_DICDI	P24639 dictyosteli
38	116	5.8	504	1	OCLN_CHICK	Q91049 gallus gall
39	116	5.8	609	1	YSW1_YEAST	P38280 saccharomyc
40	116	5.8	1464	1	CA11_HUMAN	P02452 homo sapien
41	116	5.8	2167	1	SHK1_RAT	Q9wv48 rattus norv
42	115.5	5.8	261	1	PRP2_MOUSE	P05142 mus musculu
43	115.5	5.8	475	1	S3A2_MOUSE	Q62203 mus musculu
44	115	5.7	240	1	PRA_MYCTU	O53426 mycobacteri
45	115	5.7	467	1	CBPA_DICDI	P35085 dictyosteli

ALIGNMENTS

RESULT 1
T101_HUMAN
ID T101_HUMAN STANDARD; PRT; 390 AA.
AC Q99816;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tumor susceptibility gene 101 protein.
GN TSG101.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=97148696; PubMed=9019400;
RA Li L., Li X., Francke U., Cohen S.N.;
RT "The TSG101 tumor susceptibility gene is located in chromosome 11 band
RT p15 and is mutated in human breast cancer.";
RL Cell 88:143-154(1997).
RN [2]
RP ERRATUM.
RX MEDLINE=99080983; PubMed=9867424;
RA Li L., Francke U., Cohen S.N.;
RL Cell 93:660-660(1998).
RN [3]
RP ALTERNATIVE SPLICING (ISOFORM 2).
RX MEDLINE=98031856; PubMed=9366528;
RA Gayther S.A., Barski P., Batley S.J., Li L., de Foy K.A., Cohen S.N.,
RA Ponder B.A., Caldas C.;
RT "Aberrant splicing of the TSG101 and FHIT genes occurs frequently in
RT multiple malignancies and in normal tissues and mimics alterations
RT previously described in tumours.";
RL Oncogene 15:2119-2126(1997).
RN [4]
RP ALTERNATIVE SPLICING.
RX MEDLINE=97384943; PubMed=9242438;
RA Lee M.P., Feinberg A.P.;
RT "Aberrant splicing but not mutations of TSG101 in human breast
RT cancer.";
RL Cancer Res. 57:3131-3134(1997).
RN [5]
RP ALTERNATIVE SPLICING.
RX MEDLINE=99054675; PubMed=9840940;
RA Wagner K.-U., Dierisseau P., Rucker E.B. III, Robinson G.W.,
RA Hennighausen L.;
RT "Genomic architecture and transcriptional activation of the mouse and
RT human tumor susceptibility gene TSG101: common types of shorter
RT transcripts are true alternative splice variants.";
RL Oncogene 17:2761-2770(1998).
RN [6]
RP INTERACTION WITH DMAP1.
RX MEDLINE=20347709; PubMed=10888872;
RA Rountree M.R., Bachman K.E., Baylin S.B.;
RT "DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at

```
RT   replication foci.";
RL   Nat. Genet. 25:269-277(2000).
CC   -!- FUNCTION: MAY BE INVOLVED IN CELL GROWTH AND DIFFERENTIATION AND
CC       ACT AS A NEGATIVE GROWTH REGULATOR.
CC   -!- SUBUNIT: INTERACTS WITH STATHMIN AND DMAP1.
CC   -!- SUBCELLULAR LOCATION: MAINLY CYTOPLASMIC. DEPENDING ON THE STAGE
CC       OF THE CELL CYCLE, DETECTED IN THE NUCLEUS.
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; 1 (SHOWN HERE) AND 2;
CC       ARE PRODUCED BY ALTERNATIVE SPLICING. ISOFORM 2 IS DETECTED IN
CC       NORMAL AS WELL AS CANCER TISSUES. SEVERAL SHORTER ISOFORMS ARE
CC       DETECTED IN PRIMARY BREAST CANCERS AND OTHER TUMORS.
CC   -!- TISSUE SPECIFICITY: HEART, BRAIN, PLACENTA, LUNG, LIVER, SKELETAL,
CC       KIDNEY AND PANCREAS.
CC   -!- DOMAIN: THE COILED COIL DOMAIN MAY INTERACT WITH STATHMIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U82130; AAC52083.1; -.
DR   Genew; HGNC:15971; TSG101.
DR   MIM; 601387; -.
DR   InterPro; IPR000608; UBQ_conjugat.
DR   SMART; SM00212; UBCc; 1.
KW   Growth regulation; Coiled coil; Alternative splicing.
FT   DOMAIN      235    316       COILED COIL (POTENTIAL).
FT   VARSPLIC     15    119       MISSING (IN ISOFORM 2).
SQ   SEQUENCE   390 AA;  43944 MW;  ADD6912FC22DF162 CRC64;

  Query Match             100.0%;  Score 2002;  DB 1;  Length 390;
  Best Local Similarity   100.0%;  Pred. No. 4e-105;
  Matches  380;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         11 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 70

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         71 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 130

Qy        121 MIVVFGDEPPVFSRPISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCPY 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        131 MIVVFGDEPPVFSRPISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCPY 190

Qy        181 PPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRAQ 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        191 PPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRAQ 250

Qy        241 AELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQSE 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        251 AELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQSE 310

Qy        301 NNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRKQ 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        311 NNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRKQ 370

Qy        361 FQLRALMQKARKTAGLSDLY 380
              ||||||||||||||||||||
Db        371 FQLRALMQKARKTAGLSDLY 390


RESULT 2
T101_MOUSE
ID   T101_MOUSE     STANDARD;      PRT;   391 AA.
AC   Q61187;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Tumor susceptibility gene 101 protein.
GN   TSG101.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Fibroblast;
RX   MEDLINE=96201522; PubMed=8616888;
RA   Li L., Cohen S.N.;
RT   "Tsg101: a novel tumor susceptibility gene isolated by controlled
RT   homozygous functional knockout of allelic loci in mammalian cells.";
RL   Cell 85:319-329(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/SvJ; TISSUE=Mammary gland;
RX   MEDLINE=99054675; PubMed=9840940;
RA   Wagner K.-U., Dierisseau P., Rucker E.B. III, Robinson G.W.,
RA   Hennighausen L.;
RT   "Genomic architecture and transcriptional activation of the mouse and
RT   human tumor susceptibility gene TSG101: common types of shorter
RT   transcripts are true alternative splice variants.";
RL   Oncogene 17:2761-2770(1998).
RN   [3]
RP   INTERACTION WITH DMAP1.
RX   MEDLINE=20347709; PubMed=10888872;
RA   Rountree M.R., Bachman K.E., Baylin S.B.;
RT   "DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at
RT   replication foci.";
RL   Nat. Genet. 25:269-277(2000).
CC   -!- FUNCTION: MAY BE INVOLVED IN CELL GROWTH AND DIFFERENTIATION AND
CC       ACT AS A NEGATIVE GROWTH REGULATOR.
CC   -!- SUBUNIT: INTERACTS WITH STATHMIN AND DMAP1.
CC   -!- SUBCELLULAR LOCATION: MAINLY CYTOPLASMIC. DEPENDING ON THE STAGE
CC       OF THE CELL CYCLE, DETECTED IN THE NUCLEUS.
CC   -!- TISSUE SPECIFICITY: UBIQUITOUS. HIGHER EXPRESSION IN BRAIN AND
CC       MAMMARY GLAND. LOWER EXPRESSION IN LIVER AND TUMORAL TISSUES.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED AT ALL STAGES OF MAMMARY GLAND
CC       DEVELOPMENT, BUT AT LOWER RATE AT EARLY AND MID PREGNANCY.
CC       EXPRESSED IN 1-CELL AND 2-CELL STAGE EMBRYOS.
CC   -!- DOMAIN: THE COILED COIL DOMAIN MAY INTERACT WITH STATHMIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U52945; AAC53586.1; -.
DR   EMBL; AF060868; AAC83576.1; -.
DR   MGD; MGI:106581; tsg101.
DR   InterPro; IPR000608; UBQ_conjugat.
DR   SMART; SM00212; UBCc; 1.
KW   Growth regulation; Coiled coil.
FT   DOMAIN      237    317       COILED COIL (POTENTIAL).
SQ   SEQUENCE   391 AA;  44123 MW;  79558EC535093492 CRC64;

  Query Match             94.9%;  Score 1900.5;  DB 1;  Length 391;
  Best Local Similarity   94.5%;  Pred. No. 1.8e-99;
  Matches  360;  Conservative   11;  Mismatches    9;  Indels    1;  Gaps    1;

Qy          1 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 60
              |:||||||||||:||||| :||||||||||||||||||:|||||||| |||| |||
Db         11 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNI 70

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120
              ||||||||||||||||||||||||||||||||||||||||||:||||:|:|| |||:
Db         71 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQI 130

Qy        121 MIVVFGDEPPVFSRP-ISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCP 179
```

```
                ||||:||:||||||||| :|||||||| |||||||||||||| ||| ||||||||||||||||
Db        131 MIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCP 190

Qy        180 YPPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRA 239
                ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        191 YPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGA 250

Qy        240 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 299
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        251 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 310

Qy        300 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 359
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        311 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 370

Qy        360 QFQLRALMQKARKTAGLSDLY 380
                |||||||||||||||||||||
Db        371 QFQLRALMQKARKTAGLSDLY 391
```

```
RESULT 3
ANX7_HUMAN
ID   ANX7_HUMAN     STANDARD;      PRT;   466 AA.
AC   P20073;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Annexin A7 (Annexin VII) (Synexin).
GN   ANXA7 OR ANX7.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=89264510; PubMed=2542947;
RA   Burns A.L., Magendzo K., Shirvan A., Srivastava M., Rojas E.,
RA   Alijani M.R., Pollard H.B.;
RT   "Calcium channel activity of purified human synexin and structure of
RT   the human synexin gene.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:3798-3802(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RA   Strausberg R.;
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: CALCIUM/PHOSPHOLIPID-BINDING PROTEIN WHICH PROMOTES
CC       MEMBRANE FUSION AND IS INVOLVED IN EXOCYTOSIS.
CC   -!- DOMAIN: A pair of annexin repeats may form one binding site for
CC       calcium and phospholipid.
CC   -!- SIMILARITY: BELONGS TO THE ANNEXIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 4 ANNEXIN REPEATS.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; J04543; AAA36616.1; -.
DR   EMBL; BC002632; AAH02632.1; -.
DR   PIR; A32554; LUHU7.
DR   HSSP; P26256; 1DM5.
DR   Genew; HGNC:545; ANXA7.
DR   MIM; 186360; -.
DR   InterPro; IPR001464; Annexin.
DR   Pfam; PF00191; annexin; 4.
DR   PRINTS; PR00196; ANNEXIN.
DR   ProDom; PD000143; Annexin; 4.
DR   SMART; SM00335; ANX; 4.
DR   PROSITE; PS00223; ANNEXIN; 4.
KW   Annexin; Calcium/phospholipid-binding; Repeat.
FT   DOMAIN        1    143       REPEAT-RICH REGION.
FT   REPEAT      172    232       ANNEXIN 1.
FT   REPEAT      244    304       ANNEXIN 2.
FT   REPEAT      327    387       ANNEXIN 3.
FT   REPEAT      403    463       ANNEXIN 4.
FT   DOMAIN        5     20       3 X 5 AA TANDEM REPEATS OF G-Y-P-P-X.
FT   REPEAT        5      9       1.
FT   REPEAT       10     14       2.
FT   REPEAT       16     20       3.
SQ   SEQUENCE   466 AA;  50315 MW;  09A6760729D45FCD CRC64;

  Query Match             8.0%;  Score 160;  DB 1;  Length 466;
  Best Local Similarity   21.6%;  Pred. No. 0.029;
  Matches   84;  Conservative   37;  Mismatches   90;  Indels  178;  Gaps   17;

Qy        139 SYPPYQATG-PPNTSYMPG------MPGGISPYPSGYP---------------------- 169
                ||| |  || ||   | |        | |  |||||:|
Db          2 SYPGYPPTGYPPFPGYPPAGQESSFPPSGQYPYPSGFPPMGGGAYPQVPSSGYPGAGGYP 61

Qy        170 -----PNPSGYPGCPYPPGGP-YPATTSSQ-------------YPSQPPVTTVG------ 204
                    | | |||| | | | | ||     |              || |||  : |
Db         62 APGGYPAPGGYPGAPQPGGAPSYPGVPPGQGFGVPPGGAGFSGYP-QPPSQSYGGGPAQV 120

Qy        205 ----------------------PSRDGTISE-----------DTIRASLI--------- 221
                                     ||:  |:::           | || : |
Db        121 PLPGGFPGGQMPSQYPGGQPTYPSQPATVTQVTQGTIRPAANFDAIRDAEILRKAMKGFG 180

Qy        222 ---SAVSDKLRWRMKEEMDRAQAELNAL--KRTEEDLKK---------------------- 255
                  |: | :  |  ::  : :|     |   :|||
Db        181 TDEQAIVDVVANRSNDQRQKIKAAFKTSYGKDLIKDLKSELSGNMEELILALFMPPTYYD 240

Qy        256 -----------GHQK---LEEMVTRLDQEVAEV------DKNIELLKKKDEELSSALEKM 295
                        | |:   :| : || :||: |:      :   :| |    : |   |::
Db        241 AWSLRKAMQGAGTQERVLIEILCTRTNQEIREIVRCYQSEFGRDLEKDIRSDTSGHFERL 300

Qy        296 ----------ENQSENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVID 345
                         |||| |: :                  |:|:|    ::  ||    | :
Db        301 LVSMCQGNRDENQSINHQM-----------------AQEDA--QRLYQAGE----GRLG 336

Qy        346 LDVFLKHVRLLSRKQFQLRALMQKARKTA 374
                |    :: | :|   |||| |:   : |
Db        337 TDESCFNMILATRSFPQLRATMEAYSRMA 365
```

```
RESULT 4
YCA8_YEAST
ID   YCA8_YEAST     STANDARD;      PRT;   212 AA.
AC   P25604; P87010;
DT   01-MAY-1992 (Rel. 22, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical 24.1 kDa protein in ILV6-CWH36 intergenic region.
GN   YCL008C OR YCL8C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Oliver S.G., Anwar R., Brown A., Gent M.E., Indge K.J., James C.M.,
RA   Stateva L.I.;
RL   Submitted (MAR-1992) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   REVISIONS.
RA   Gromadka R.;
RL   Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X59720; CAA42351.1; -.
DR   PIR; S19410; S19410.
DR   SGD; S0000514; STP22.
DR   InterPro; IPR000608; UBQ_conjugat.
DR   SMART; SM00212; UBCc; 1.
DR   PROSITE; PS00183; UBIQUITIN_CONJUGAT_1; FALSE_NEG.
DR   PROSITE; PS50127; UBIQUITIN_CONJUGAT_2; 1.
KW   Hypothetical protein; Ubl conjugation pathway; Ligase.
SQ   SEQUENCE   212 AA;  24086 MW;  AE5309BB58A2B2B8 CRC64;

  Query Match             7.3%;  Score 145.5;  DB 1;  Length 212;
  Best Local Similarity   24.5%;  Pred. No. 0.074;
  Matches   58;  Conservative   38;  Mismatches  100;  Indels   41;  Gaps   11;

Qy         64 LWLLDTYPYNPPIC------FVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGL 117
              :|:   ||  ||          |   | | ::   :::|:|| | || || |     :|: :
Db          1 MWVPSMYPVKPPFISINLENFDMNTISSSLPIQEYIDSNGWIALPILHCWDPAAMNLIMV 60

Qy        118 IQVMIVVFGDEPPVFSRPISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPG 177
              :| :: :   |||     |      |  |      ||  :|  |       |   || |
Db         61 VQELMSLL-HEPPQDQAPSLPPKPNTQLQQEQNTPPLPPKPKS----PHLKPPLP----- 110

Qy        178 CPYPPGGPYPATTSSQYPSQPPVTTVGPSRD-------GTISEDTIRASLISAVSDKLRW 230
                ||  | ||: :          | : |:           |:  :  |   :  |:||:
Db        111 ---PPPPPQPASNALDLMDMDN-TDISPTNHHEMLQNLQTVVNELYRED-VDYVADKILT 165

Qy        231 R---MKEEMDRAQAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKK 284
              |   |:| : |   |: |       : | | :  |    ::| :  ::  |::| :|
Db        166 RQTVMQESIARFH-EIIA-------IDKNHLRAVEQA--IEQTMHSLNAQIDVLNRK 212


RESULT 5
ANXB_HUMAN
ID   ANXB_HUMAN     STANDARD;      PRT;   505 AA.
AC   P50995;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Annexin A11 (Annexin XI) (Calcyclin-associated annexin 50) (CAP-50)
DE   (56 kDa autoantigen).
GN   ANXA11 OR ANX11.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Teratocarcinoma;
RX   MEDLINE=94140847; PubMed=7508441;
RA   Misaki Y., Pruijn G.J., van der Kemp A.W., van Venrooij W.J.;
RT   "The 56K autoantigen is identical to human annexin XI.";
RL   J. Biol. Chem. 269:4240-4246(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20469408; PubMed=11013079;
RA   Bances P., Fernandez M.R., Rodriguez-Garcia M.I., Morgan R.O.,
RA   Fernandez M.P.;
RT   "Annexin A11 (ANXA11) gene structure as the progenitor of paralogous
RT   annexins and source of orthologous cDNA isoforms.";
RL   Genomics 69:95-103(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lymph;
RA   Strausberg R.;
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC AND POSSIBLY NUCLEAR.
CC   -!- DOMAIN: A pair of annexin repeats may form one binding site for
CC       calcium and phospholipid.
CC   -!- DISEASE: ANTIBODIES AGAINST ANXA11 ARE PRESENT IN SERA FROM
CC       PATIENTS WITH VARIOUS AUTOIMMUNE DISEASES, PREDOMINANTLY IN SERA
CC       FROM PATIENTS WITH RHEUMATOID ARTHRITIS, SYSTEMIC LUPUS
CC       ERYTHEMATOSUS, OR SJOGREN'S SYNDROME.
CC   -!- SIMILARITY: BELONGS TO THE ANNEXIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 4 ANNEXIN REPEATS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L19605; AAA19734.1; -.
DR   EMBL; AJ278463; CAB94995.1; -.
DR   EMBL; AJ278464; CAB94996.1; -.
DR   EMBL; AJ278465; CAB94997.1; -.
DR   EMBL; BC007564; AAH07564.1; -.
DR   HSSP; P13214; 1AOW.
DR   Genew; HGNC:535; ANXA11.
DR   MIM; 602572; -.
DR   InterPro; IPR001464; Annexin.
DR   Pfam; PF00191; annexin; 4.
DR   PRINTS; PR00196; ANNEXIN.
DR   ProDom; PD000143; Annexin; 4.
DR   SMART; SM00335; ANX; 4.
DR   PROSITE; PS00223; ANNEXIN; 4.
KW   Annexin; Calcium/phospholipid-binding; Repeat; Polymorphism.
FT   REPEAT      209    269       ANNEXIN 1.
FT   REPEAT      281    341       ANNEXIN 2.
FT   REPEAT      365    425       ANNEXIN 3.
FT   REPEAT      440    500       ANNEXIN 4.
FT   VARIANT     230    230       R -> C (IN DBSNP:1049550).
FT                                /FTId=VAR_012006.
FT   VARIANT     457    457       I -> V (IN DBSNP:1802932).
FT                                /FTId=VAR_012007.
SQ   SEQUENCE   505 AA;  54389 MW;  4ADCAC8F270BFEE4 CRC64;

  Query Match             7.1%;  Score 142;  DB 1;  Length 505;
  Best Local Similarity   28.0%;  Pred. No. 0.32;
  Matches   63;  Conservative   19;  Mismatches   91;  Indels   52;  Gaps   10;

Qy        111 QSDLLGLIQVMIVVFG--DEPPVFSRPISASYPPYQ--ATGPPNTSYMPGMPGGISPYPS 166
              |  | |:   |   ||  : | ::     | |||      | | ::  |  | |: | |
Db         51 QDYLSGMAANMSGTFGGANMPNLYPGAPGAGYPPVPPGGFGQPPSAQQPVPPYGMYPPPG 110

Qy        167 GYPPN--PS--GYPGCPYP----------PGGPYPATTSSQYPSQPPVTTVG-----PSR 207
              | ||:  ||   ||| | |          | | ||      || ||||   |     ||
Db        111 GNPPSRMPSYPPYPGAPVPGQPMPPPGQQPPGAYPGQPPVTYPGQPPVPLPGQQQPVPSY 170

Qy        208 DGTISEDTI----------RASLISAV-------SDKLRWRMKEEMDRAQAELNALKRT 249
              |     |:          | ::  |        :: ||  ||      || :: |
Db        171 PGYPGSGTVTPAVPPTQFGSRGTITDAPGFDPLRDAEVLRKAMKGFGTDEQAIIDCLGSR 230

Qy        250 EEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEK 294
                ):  | |    |   :         :|:|    |||   ||
Db        231 SN--KQRQQILLSFKTAYGK---------DLIKDLKSELSGNFEK 264


RESULT 6
ANXB_MOUSE
ID   ANXB_MOUSE     STANDARD;      PRT;   503 AA.
AC   P97384;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Annexin A11 (Annexin XI) (Calcyclin-associated annexin 50) (CAP-50).
GN   ANXA11 OR ANX11.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97092887; PubMed=8938449;
RA   Fernandez M.P., Jenkins N.A., Gilbert D.J., Copeland N.G.,
RA   Morgan R.O.;
RT   "Sequence and chromosomal localization of mouse annexin XI.";
RL   Genomics 37:366-374(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/SvJ;
RX   MEDLINE=20469408; PubMed=11013079;
RA   Bances P., Fernandez M.R., Rodriguez-Garcia M.I., Morgan R.O.,
RA   Fernandez M.P.;
RT   "Annexin A11 (ANXA11) gene structure as the progenitor of paralogous
RT   annexins and source of orthologous cDNA isoforms.";
RL   Genomics 69:95-103(2000).
CC   -!- FUNCTION: BINDS SPECIFICALLY TO CALCYCLIN IN A CALCIUM-DEPENDENT
CC       MANNER.
CC   -!- DOMAIN: A pair of annexin repeats may form one binding site for
CC       calcium and phospholipid.
CC   -!- SIMILARITY: BELONGS TO THE ANNEXIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 4 ANNEXIN REPEATS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U65986; AAB42012.1; -.
DR   EMBL; AJ289760; CAB94770.1; -.
DR   EMBL; AJ289761; CAB94770.1; JOINED.
DR   EMBL; AJ289762; CAB94770.1; JOINED.
DR   EMBL; AJ289763; CAB94770.1; JOINED.
DR   EMBL; AJ289764; CAB94770.1; JOINED.
DR   EMBL; AJ289765; CAB94770.1; JOINED.
DR   EMBL; AJ289766; CAB94770.1; JOINED.
DR   EMBL; AJ289767; CAB94770.1; JOINED.
DR   EMBL; AJ289768; CAB94770.1; JOINED.
DR   EMBL; AJ289769; CAB94770.1; JOINED.
DR   HSSP; P13214; 1ANN.
DR   SWISS-2DPAGE; P97384; MOUSE.
DR   MGD; MGI:108481; Anxa11.
DR   InterPro; IPR001464; Annexin.
DR   Pfam; PF00191; annexin; 4.
DR   PRINTS; PR00196; ANNEXIN.
DR   ProDom; PD000143; Annexin; 4.
DR   SMART; SM00335; ANX; 4.
DR   PROSITE; PS00223; ANNEXIN; 4.
KW   Annexin; Calcium/phospholipid-binding; Repeat.
FT   REPEAT      207    267       ANNEXIN 1.
FT   REPEAT      279    339       ANNEXIN 2.
FT   REPEAT      363    423       ANNEXIN 3.
FT   REPEAT      438    498       ANNEXIN 4.
SQ   SEQUENCE   503 AA;  54111 MW;  424B1345E0F4EC8A CRC64;

  Query Match             7.1%;  Score 141.5;  DB 1;  Length 503;
  Best Local Similarity   26.0%;  Pred. No. 0.34;
  Matches   76;  Conservative   29;  Mismatches   96;  Indels   91;  Gaps   16;

Qy        125 FGDEPPVFSRPISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCPYPPGG 184
              ||  ||   :|:    |||    ||  :  ||||     ||: ||  |   || | || |
Db         89 FGQPPPA-QQPV----PPYGMYPPPGGNPPPGMPS----YPA-YPGAP--VPGQPMPPTG 136

Qy        185 PYPATTSSQYPSQPPVTTVG----------PSRDGTISEDTIRASL----------IS 222
                  |      || |||:|  |           ||  |     ||  ::          |:
Db        137 QQP---PGAYPGQPPMTYPGQSPMPPPGQQPVPSYPGYSGSSTITPAVPPAQFGNRGTIT 193

Qy        223 AVS--------DKLRWRMKEEMDRAQAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEV 274
              | |        : ||  ||      || :: |       |:  | |    |   :
Db        194 AASGFDPLRDAEVLRKAMKGFGTDEQAIIDCLGSRSN--KQRQQILLSFKTAYGK----- 246

Qy        275 DKNIELLKKKDEELSSALEKMENQSENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFY 334
                  :|:|    |||   |                        | || |   :  :   ::
Db        247 ----DLIKDLKSELSGNFE---------------------KTILALM--KTPVLFDVYE 278

Qy        335 LGEALRRGVID----LDVFL----KHVRLLSRKQFQLRALMQKARKTAGLSD 378
                : ||::    |    :::|     :|:| |||     :   ||  : |  ||
Db        279 IKEAIKGAGTDEACLIEIFASRSNEHIRELSR---AYKTEFQKTLEEAIRSD 327


RESULT 7
ANX7_MOUSE
ID   ANX7_MOUSE     STANDARD;      PRT;   463 AA.
AC   Q07076;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Annexin A7 (Annexin VII) (Synexin).
GN   ANXA7 OR ANX7.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93168121; PubMed=7916616;
RA   Zhang-Keck Z.Y., Burns A.L., Pollard H.B.;
RT   "Mouse synexin (annexin VII) polymorphisms and a phylogenetic
RT   comparison with other synexins.";
RL   Biochem. J. 289:735-741(1993).
CC   -!- FUNCTION: CALCIUM/PHOSPHOLIPID-BINDING PROTEIN WHICH PROMOTES
CC       MEMBRANE FUSION AND IS INVOLVED IN EXOCYTOSIS.
CC   -!- DOMAIN: A pair of annexin repeats may form one binding site for
CC       calcium and phospholipid.
CC   -!- SIMILARITY: BELONGS TO THE ANNEXIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 4 ANNEXIN REPEATS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L13129; AAA37238.1; -.
DR   HSSP; P26256; 1DM5.
DR   MGD; MGI:88031; Anxa7.
DR   InterPro; IPR001464; Annexin.
DR   Pfam; PF00191; annexin; 4.
DR   PRINTS; PR00196; ANNEXIN.
DR   ProDom; PD000143; Annexin; 4.
DR   SMART; SM00335; ANX; 4.
DR   PROSITE; PS00223; ANNEXIN; 4.
KW   Annexin; Calcium/phospholipid-binding; Repeat.
FT   DOMAIN        1    143       REPEAT-RICH REGION.
FT   REPEAT      169    229       ANNEXIN 1.
FT   REPEAT      241    301       ANNEXIN 2.
FT   REPEAT      324    384       ANNEXIN 3.
FT   REPEAT      400    460       ANNEXIN 4.
FT   DOMAIN        5     20       3 X 5 AA TANDEM REPEATS OF G-Y-P-P-X.
FT   REPEAT        5      9       1.
FT   REPEAT       10     14       2.
FT   REPEAT       16     20       3.
SQ   SEQUENCE   463 AA;  49939 MW;  50F7B20FD48EBCD5 CRC64;

  Query Match             7.0%;  Score 141;  DB 1;  Length 463;
  Best Local Similarity   22.1%;  Pred. No. 0.33;
  Matches   73;  Conservative   42;  Mismatches  108;  Indels  108;  Gaps   15;

Qy        139 SYPPYQATGPPNTSYMP---------GMPGGISP-----YPSGY----PPNPSGYPGCPY 180
              :|||  : | |     |          | || :||     || |     ||  :|: | |
Db         45 AYPPAPSGGYPGAGGYPAPGGYPAPGGYPGALSPGGPPAYPGGQGFGAPPGGAGFSGYPQ 104
```

```
Qy        181 PP-----GGP-----------------YPATTSSQYPSQPPVTTVGPSRDGTI----SED 214
              ||     |||                 ||    : |||||   | |    |||    : |
Db        105 PPAQSYGGGPAQVPVPGGFPGGQMPSQYPG-GQAPYPSQPASMTQG--TQGTILPASNFD 161

Qy        215 TIRASLI-----------SAVSDKLRWRMKEEMDRAQAELNAL--KRTEEDLKK----- 255
               :| : |           |: | : . |  ::  : :|    :  |   :|||
Db        162 AMRDAEILRKAMKGFGTDEQAIVDVVSNRSNDQRQQIKAAFKTMYGKDLIKDLKSELSGN 221

Qy        256 -------------------------GHQK---LEEMVTRLDQEVAEVDKNIELLKKKD 285
                                        | |:   :| : || :||: :: :  :|
Db        222 MEELILALFMPSTYYDAWSLRKAMQGAGTQERVLIEILCTRTNQEIRDIVRCYQL----- 276

Qy        286 EELSSALEKMENQSENNDIDEVIIPTAPLYKQILNLYAEENAIEDT--IFYLGEALRRGV 343
              |    |||      :   : \::      :        : | ||   ::  ||    |
Db        277 -EFGRDLEKDIRSDTSGHFERLLVSMCQANRDERQSVNHQMAQEDAQRLYQAGE----GR 331

Qy        344 IDLDVFLKHVRLLSRKQFQLRALMQKARKTA 374
              :  |    :: | :|   ||:| |:   : |
Db        332 LGTDESCFNMILATRSFPQLKATMEAYSRMA 362


RESULT 8
ANXB_RABIT
ID   ANXB_RABIT     STANDARD;      PRT;   503 AA.
AC   P33477;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Annexin A11 (Annexin XI) (Calcyclin-associated annexin 50) (CAP-50).
GN   ANXA11 OR ANX11.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Lung;
RX   MEDLINE=92378579; PubMed=1380798;
RA   Tokumitsu H., Mizutani A., Muramatsu M.-A., Yokota T., Arai K.-I.,
RA   Hidaka H.;
RT   "Molecular cloning of rabbit CAP-50, a calcyclin-associated annexin
RT   protein.";
RL   Biochem. Biophys. Res. Commun. 186:1227-1235(1992).
CC   -!- FUNCTION: BINDS SPECIFICALLY TO CALCYCLIN IN A CALCIUM-DEPENDENT
CC       MANNER.
CC   -!- DOMAIN: A pair of annexin repeats may form one binding site for
CC       calcium and phospholipid.
CC   -!- SIMILARITY: BELONGS TO THE ANNEXIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 4 ANNEXIN REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D10883; BAA01705.1; -.
DR   PIR; JH0694; LURB11.
DR   HSSP; P13214; 1ANN.
DR   InterPro; IPR001464; Annexin.
DR   Pfam; PF00191; annexin; 4.
DR   PRINTS; PR00196; ANNEXIN.
DR   ProDom; PD000143; Annexin; 4.
DR   SMART; SM00335; ANX; 4.
DR   PROSITE; PS00223; ANNEXIN; 4.
KW   Annexin; Calcium/phospholipid-binding; Repeat.
FT   REPEAT      207    267       ANNEXIN 1.
FT   REPEAT      279    339       ANNEXIN 2.
FT   REPEAT      363    423       ANNEXIN 3.
FT   REPEAT      438    498       ANNEXIN 4.
SQ   SEQUENCE   503 AA;  54034 MW;  44C15F290770AC9F CRC64;

  Query Match             6.8%;  Score 136.5;  DB 1;  Length 503;
  Best Local Similarity   24.1%;  Pred. No. 0.65;
  Matches   81;  Conservative   39;  Mismatches  117;  Indels   99;  Gaps   18;

Qy        126 GDEPPV----FSRPISASYPPYQATGPPNTSYMPGMPGGISPYPSG---YPPNPSG-YPG 177
              |  |||    | :|     || | : ||   | |  ||| :| |||   ||| |   ||
Db         79 GGYPPVPPGGFGQP-----PPTQPSVPPYGVYPP--PGG-NP-PSGVPSYPPFPGAPVPG 129

Qy        178 CPYP-----PGGPYPATTSSQYPSQPPVTTVG----PSRDGTISEDTI----------R 217
               | |     | ||||      || | ||   |    ||  |    |:          |
Db        130 QPMPPPGHQPPGPYPGQLPVTYPGQSPVPPPGQQPMPSYPGYPGSGTVTPAVPPVQFGNR 189

Qy        218 ASLISAV-------SDKLRWRMKEEMDRAQAELNAL-----KRTEEDLKK-----GHQKL 260
              ::  |        :: ||  ||      || :: |     |: :: |       |   :
Db        190 GTITDASGFDPLRDAEVLRKAMKGFGTDEQAIIDCLGSRSNKQRQQILLSFKTAYGKDLI 249

Qy        261 EEMVTRLD-----------------QEVAEVDKN--------IELLKKKDEELSSALE 293
              ::: : |                 |: |  |          ||:|  :  |    |
Db        250 KDLKSELSGNFEKTILALMKTPILFDAYEIKEAIKGAGTDEACLIEILASRSNEHIRELN 309

Qy        294 KMENQSENNDIDEVI-IPTAPLYKQILNLYAEENAIEDT-------------IFYLGEAL 339
              |          ::| |   |:  ::::|   :: |  | |             ::  ||
Db        310 KAYKTEFKKTLEEAIRSDTSGHFQRLLISLSQGNRDESTNVDMSLVQRDVQELYAAGE-- 367

Qy        340 RRGVIDLDVFLKHVRLLSRKQFQLRALMQKARKTAG 375
                :  |    :  | || :  | |:  : ::  |
Db        368 --NRLGTDESKFNAVLCSRSRAHLVAVFNEYQRMTG 401


RESULT 9
FOSB_HUMAN
ID   FOSB_HUMAN     STANDARD;      PRT;   338 AA.
AC   P53539;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Protein fosB (G0/G1 switch regulatory protein 3).
GN   FOSB OR G0S3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93250970; PubMed=1301997;
RA   Martin-Gallardo A., McCombie W.R., Gocayne J.D., Fitzgerald M.G.,
RA   Wallace S., Lee B.M., Lamerdin J., Trapp S., Kelley J.M.,
RA   Liu L.-I., Dubnick M., Johnston-Dow L.A., Kerlavage A.R.,
RA   de Jong P., Carrano A., Fields C., Venter J.C.;
RT   "Automated DNA sequencing and analysis of 106 kilobases from human
RT   chromosome 19q13.3.";
RL   Nat. Genet. 1:34-39(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Blood;
RX   MEDLINE=97138090; PubMed=8985116;
RA   Heximer S.P., Cristillo A.D., Russell L., Forsdyke D.R.;
RT   "Sequence analysis and expression in cultured lymphocytes of the
RT   human FOSB gene (G0S3).";
RL   DNA Cell Biol. 15:1025-1038(1996).
CC   -!- FUNCTION: FOSB INTERACTS WITH JUN PROTEINS ENHANCING THEIR DNA
CC       BINDING ACTIVITY.
CC   -!- SUBUNIT: HETERODIMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: BELONGS TO THE BZIP FAMILY. FOS SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L49169; AAB53946.1; -.
DR   HSSP; P01100; 1FOS.
DR   Genew; HGNC:3797; FOSB.
DR   MIM; 164772; -.
DR   InterPro; IPR000837; Leuzip_Fos.
DR   InterPro; IPR004827; TF_bZIP.
DR   Pfam; PF00170; bZIP; 1.
DR   PRINTS; PR00042; LEUZIPPRFOS.
DR   SMART; SM00338; BRLZ; 1.
DR   PROSITE; PS00036; BZIP_BASIC; 1.
KW   Nuclear protein; DNA-binding.
FT   DNA_BIND     161    179       BASIC MOTIF.
FT   DOMAIN       183    211       LEUCINE-ZIPPER.
FT   CONFLICT     338    338       L -> R (IN REF. 2).
SQ   SEQUENCE   338 AA;  35927 MW;  DDFF827C5047850F CRC64;

  Query Match             6.5%;  Score 130.5;  DB 1;  Length 338;
  Best Local Similarity   28.3%;  Pred. No. 0.88;
  Matches   51;  Conservative   18;  Mismatches   68;  Indels   43;  Gaps    7;

Qy        113 DLLGLIQVMIV---------VFGDEPPVFSRPISASYPPYQATGPPNTSY-MPGMPGGIS 162
              ||  |:|  ::           :|||          ||     | |||  ||| |  |
Db         66 DLQWLVQPTLISSMAQSQGQPLASQPPVVD--------PYDM---PGTSYSTPGMSGYSS 114

Qy        163 PYPSGYPPNPSGYPGCPYPPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLIS 222
                ||     || |       || ||  :   | :|   |: | :    :  :|
Db        115 GGASG-----SGGPSTSGTTSGPGPARPARARPRRPREETLTPEEE---EKRRVRRERNK 166

Qy        223 AVSDKLRWRMKEEMDRAQAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLK 282
                : | | :|  || ||| :           :|||    |: |:||: |  | |:
Db        167 LAAAKCRNRRRELTDRLQAETD-------------QLEEEKAELESEIAELQKEKERLE 212


RESULT 10
SN24_HUMAN
ID   SN24_HUMAN     STANDARD;      PRT;  1647 AA.
AC   P51532;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Possible global transcription activator SNF2L4 (SNF2-beta) (BRG-1
DE   protein) (Mitotic growth and transcription activator) (Brahma
DE   protein homolog 1).
GN   SMARCA4 OR SNF2L4 OR BRG1 OR SNF2B.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94050144; PubMed=8232556;
RA   Khavari P.A., Peterson C.L., Tamkun J.W., Mendel D.B., Crabtree G.R.;
RT   "BRG1 contains a conserved domain of the SWI2/SNF2 family necessary
RT   for normal mitotic growth and transcription.";
RL   Nature 366:170-174(1993).
RN   [2]
RP   REVISIONS.
RA   Khavari P.A., Peterson C.L., Tamkun J.W., Mendel D.B., Crabtree G.R.;
RL   Submitted (JUN-1995) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94268902; PubMed=8208605;
RA   Chiba H., Muramatsu M., Nomoto A., Kato H.;
RT   "Two human homologues of Saccharomyces cerevisiae SWI2/SNF2 and
RT   Drosophila brahma are transcriptional coactivators cooperating with
RT   the estrogen receptor and the retinoic acid receptor.";
RL   Nucleic Acids Res. 22:1815-1820(1994).
RN   [4]
RP   SEQUENCE OF 814-1474 FROM N.A.
RA   Lamerdin J.E., McCready P.M., Skowronski E., Viswanathan V.,
RA   Burkhart-Schultz K., Gordon L., Dias J., Brower A., Stilwagen S.,
RA   Phan H., Velasco N., Do L., Regala W., Terry A., Garnes J.,
RA   Danganan L., Erler A., Christensen M., Georgescu A., Avila J., Liu S.,
RA   Attix C., Andreise T., Trankheim M., Amico-Keller G., Coefield J.,
RA   Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Kronmiller B.,
RA   Arellano A., Sanders C., Ow D., Nolan M., Trong S., Kobayashi A.,
RA   Olsen A.S., Carrano A.V.;
RL   Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: TRANSCRIPTIONAL COACTIVATOR COOPERATING WITH NUCLEAR
CC       HORMONE RECEPTORS TO POTENTIATE TRANSCRIPTIONAL ACTIVATION.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U29175; AAB40977.1; -.
DR   EMBL; D26156; BAA05143.1; -.
DR   EMBL; AC006127; AAC97987.1; -.
DR   Genew; HGNC:11100; SMARCA4.
DR   MIM; 603254; -.
DR   InterPro; IPR001487; Bromodomain.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR000330; SNF2_N.
DR   Pfam; PF00176; SNF2_N; 1.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00439; bromodomain; 1.
DR   PRINTS; PR00503; BROMODOMAIN.
DR   SMART; SM00297; BROMO; 1.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   PROSITE; PS00633; BROMODOMAIN_1; 1.
DR   PROSITE; PS50014; BROMODOMAIN_2; 1.
KW   Transcription regulation; Nuclear protein; Activator; Bromodomain;
KW   ATP-binding; Helicase.
FT   DOMAIN       578    588       POLY-LYS.
FT   DOMAIN       663    672       POLY-GLU.
FT   NP_BIND      779    786       ATP (POTENTIAL).
FT   SITE         881    884       DEGH BOX.
FT   DOMAIN      1360   1364       POLY-GLU.
FT   DOMAIN      1477   1547       BROMODOMAIN.
FT   DOMAIN      1571   1584       POLY-GLU.
SQ   SEQUENCE   1647 AA;  184585 MW;  7B785E7953277F1D CRC64;

  Query Match             6.5%;  Score 130;  DB 1;  Length 1647;
  Best Local Similarity   21.7%;  Pred. No. 6.1;
  Matches   61;  Conservative   27;  Mismatches   71;  Indels  122;  Gaps   11;

Qy        126 GDEPPVFSRPISASYPPYQATGPPNTSYMPGMP-----GGISPYPSG------------ 167
              |  || :|||     |     ||  :   ||||     |   |:| |
Db        243 GPAPPNYSRPHGMGGPNMPPPGP--SGVPPGMPGQPPGGPPKPWPEGPMANAAAPTSTPQ 300

Qy        168 --YPPNPSGYPGCPYPPGGP-------YPATTSSQYPSQPP-----------VTTVGPSR 207
                 || |:| |  | ||  |        | | |   |:||           :| :   |
Db        301 KLIPPQPTGRPS-PAPPAVPPAASPVMPPQTQSPGQPAQPAPMVPLHQKQSRITPIQKPR 359

Qy        208 ----------------------------DGTISED-----TI---------------- 216
                                           |::: |     ||
Db        360 GLDPVEILQEREYRLQARIAHRIQELENLPGSLAGDLRTKATIELKALRLLNFQRQLRQE 419

Qy        217 ----------------------RASLISA-VSDKLRWRMKEEMDRA-----QAELNAL 246
                                   | ||  | :::||  : | | :|      |  ||::
Db        420 VVVCMRRDTALETALNAKAYKRSKRQSLREARITEKLEKQQKIEQERKRRQKHQEYLNSI 479

Qy        247 KRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEE 287
                :  :| |: |: :   : :| : ||    | |  :||: |
Db        480 LQHAKDFKEYHRSVTGKIQKLTKAVATYHANTEREQKKENE 520
```

```
RESULT 11
DRPL_HUMAN
ID   DRPL_HUMAN     STANDARD;      PRT;  1185 AA.
AC   P54259;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Atrophin-1 (Dentatorubral-pallidoluysian atrophy protein).
GN   DRPLA.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Cerebellum, and Brain;
RX   MEDLINE=95144175; PubMed=7842016;
RA   Nagafuchi S., Yanagisawa H., Ohsaki E., Shirayama T., Tadokoro K.,
RA   Inoue T., Yamada M.;
RT   "Structure and expression of the gene responsible for the triplet
RT   repeat disorder, dentatorubral and pallidoluysian atrophy (DRPLA).";
RL   Nat. Genet. 8:177-182(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96262314; PubMed=8965642;
RA   Margolis R.L., Li S.-H., Young W.S., Wagster M.V., Stine O.C.,
RA   Kidwai A.S., Ashworth R.G., Ross C.A.;
RT   "DRPLA gene (atrophin-1) sequence and mRNA expression in human
RT   brain.";
RL   Brain Res. Mol. Brain Res. 36:219-226(1996).
RN   [3]
RP   SEQUENCE OF 470-725 FROM N.A.
RC   TISSUE=Brain cortex;
RX   MEDLINE=93315145; PubMed=8325628;
RA   Li S.-H., McInnis M.G., Margolis R.L., Antonarakis S.E., Ross C.A.;
RT   "Novel triplet repeat containing genes in human brain: cloning,
RT   expression, and length polymorphisms.";
RL   Genomics 16:572-579(1993).
CC   -!- TISSUE SPECIFICITY: THE LEVELS ARE RELATIVELY HIGH IN THE BRAIN,
CC       OVARY, TESTIS AND PROSTATE. LOWER LEVELS ARE DETECTED IN THE
CC       LIVER, THYMUS AND LEUKOCYTES.
CC   -!- POLYMORPHISM: THE POLY-GLN REGION OF DRPLA IS HIGHLY POLYMORPHIC
CC       (7 TO 23 REPEATS) IN THE NORMAL POPULATION AND IS EXPANDED TO
CC       ABOUT 49-75 REPEATS IN DRPLA PATIENTS. LONGER EXPANSIONS RESULT IN
CC       EARLIER ONSET AND MORE SEVERE CLINICAL MANIFESTATIONS OF THE
CC       DISEASE.
CC   -!- DISEASE: DEFECTS IN DRPLA ARE THE CAUSE OF DENTATORUBRAL-
CC       PALLIDOLUYSIAN ATROPHY, AN AUTOSOMAL DOMINANT NEURODEGENERATIVE
CC       DISORDER CHARACTERIZED BY A LOSS OF NEURONS IN THE DENTATE
CC       NUCLEUS, RUBRUM, GLOGUS PALLIDUS AND LUYS'BODY. CLINICAL FEATURES
CC       ARE MYOCLONUS EPILEPSY, DEMENTIA, AND CEREBELLAR ATAXIA. ONSET OF
CC       THE DISEASE OCCURS USUALLY IN THE SECOND DECADE OF LIFE AND DEATH
CC       IN THE FOURTH.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D31840; BAA06626.1; -.
DR   EMBL; U23851; AAB50276.1; -.
DR   EMBL; L10377; -; NOT_ANNOTATED_CDS.
DR   Genew; HGNC:3033; DRPLA.
DR   MIM; 125370; -.
DR   InterPro; IPR002951; Atrophin.
DR   Pfam; PF03154; Atrophin-1; 2.
DR   PRINTS; PR01222; ATROPHIN.
KW   Triplet repeat expansion; Polymorphism.
FT   DOMAIN       73     82       SER/GLU-RICH (MIXED CHARGE).
FT   DOMAIN      302    305       POLY-PRO.
FT   DOMAIN      376    382       POLY-SER.
FT   DOMAIN      386    397       POLY-SER.
FT   DOMAIN      442    447       POLY-PRO.
FT   DOMAIN      479    483       POLY-HIS.
FT   DOMAIN      484    497       POLY-GLN.
FT   DOMAIN      504    507       POLY-PRO.
FT   DOMAIN      564    574       POLY-SER.
FT   DOMAIN      704    707       POLY-PRO.
FT   DOMAIN      802    815       ARG/ALA-RICH (MIXED CHARGE).
FT   DOMAIN      816    827       ARG/GLU-RICH (MIXED CHARGE).
FT   DOMAIN      925    934       ARG/GLU-RICH (MIXED CHARGE).
FT   CONFLICT     94     94       MISSING (IN REF. 2).
FT   CONFLICT    333    333       Y -> H (IN REF. 2).
FT   CONFLICT    339    339       M -> I (IN REF. 2).
FT   CONFLICT    541    541       P -> T (IN REF. 3).
FT   CONFLICT   1028   1028       G -> A (IN REF. 2).
SQ   SEQUENCE   1185 AA;  124785 MW;  56C306267331C005 CRC64;

  Query Match             6.3%;  Score 127;  DB 1;  Length 1185;
  Best Local Similarity   23.6%;  Pred. No. 6.1;
  Matches   64;  Conservative   32;  Mismatches   93;  Indels   82;  Gaps   13;

Qy         70 YPYNPPICFVKPTSSMTIKT----------------------GKHVDANGKI------- 99
              ||: ||:  |  ||| |: |                        ||   : |
Db        594 YPF-PPVPTV-TTSSATLSTVIATVASSPAGYKTASPPGPPPYGKRAPSPGAYKTATPPG 651

Qy        100 YLPYLHEWKHPQSDLLGLIQVMIVVFGDEPPVFSRPISASYPPYQATGPP--NTSYMPGM 157
              | |                          || |    : : | |: | ||    :: ||
Db        652 YKP----------------------GSPPSFR---TGTPPGYRGTSPPAGPGTFKPGS 684

Qy        158 PG-GISPYPSGYPPNPSGYPGCPYPP-----GGPYPATTSSQYPS---QPPVTTVGPSRD 208
              |  |  | |   |  ||| |  | ||     | |  ||   | |:   : | : | |:|
Db        685 PTVGPGPLP---PAGPSGLPSLPPPPAAPASGPPLSATQIKQEPAEEYETPESPVPPARS 741

Qy        209 GTISEDTIRASLISAVSDKLRWRMKEEMDR-----AQAELNALKRTEEDLKKGHQKLEEM 263
                :     :    :: |    |  : :||     |:::|  :      | |    | |
Db        742 PSPPPKVVDVPSHASQSA----RFNKHLDRGFNSCARSDLYFVPLEGSKLAKKRADLVEK 797

Qy        264 VTRLDQEVAEVDKNIELLKKKDEELSSALEK 294
              | |  :: |  :|  |  :::::|     |:
Db        798 VRREAEQRAREEKERERERERERERERERER 828


RESULT 12
CA11_MOUSE
ID   CA11_MOUSE     STANDARD;      PRT;  1453 AA.
AC   P11087; Q60635;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Collagen alpha 1(I) chain precursor.
GN   COL1A1 OR COLA1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB/N;
RX   MEDLINE=96033240; PubMed=8535610;
RA   Li S.W., Khillan J., Prockop D.J.;
RT   "The complete cDNA coding sequence for the mouse pro alpha 1(I) chain
RT   of type I procollagen.";
RL   Matrix Biol. 14:593-595(1995).
RN   [2]
RP   SEQUENCE OF 518-1128 FROM N.A.
RX   MEDLINE=86137403; PubMed=3841523;
RA   French B.T., Lee W.-H., Maul G.G.;
RT   "Nucleotide sequence of a cDNA clone for mouse pro alpha 1(I)
RT   collagen protein.";
RL   Gene 39:311-312(1985).
```

```
RN   [3]
RP   SEQUENCE OF 735-1130 FROM N.A.
RX   MEDLINE=83141374; PubMed=6298597;
RA   Monson J.M., Friedman J., McCarthy B.J.;
RT   "DNA sequence analysis of a mouse pro alpha 1 (I) procollagen gene:
RT   evidence for a mouse B1 element within the gene.";
RL   Mol. Cell. Biol. 2:1362-1371(1982).
RN   [4]
RP   SEQUENCE OF 735-878 AND 1005-1058 FROM N.A.
RX   MEDLINE=83157109; PubMed=6219867;
RA   Monson J.M., McCarthy B.J.;
RT   "Identification of a Balb/c mouse pro alpha 1(I) procollagen gene:
RT   evidence for insertions or deletions in gene coding sequences.";
RL   DNA 1:59-69(1981).
RN   [5]
RP   SEQUENCE OF 1442-1453 FROM N.A.
RX   MEDLINE=88124276; PubMed=3340560;
RA   Mooslehner K., Harbers K.;
RT   "Two mRNAs of mouse pro alpha 1(I) collagen gene differ in the size
RT   of the 3'-untranslated region.";
RL   Nucleic Acids Res. 16:773-773(1988).
CC   -!- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
CC       (FIBRILLAR FORMING COLLAGEN).
CC   -!- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
CC   -!- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
CC       BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
CC       HYDROXYAPATITE.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U08020; AAA88912.1; -.
DR   EMBL; X15896; CAA33904.1; -.
DR   EMBL; M14423; AAA37333.1; -.
DR   EMBL; M17491; AAA37334.1; -.
DR   EMBL; X06753; CAA29927.1; -.
DR   EMBL; K03036; AAA37332.1; -.
DR   EMBL; K03029; AAA37332.1; JOINED.
DR   EMBL; K03030; AAA37332.1; JOINED.
DR   EMBL; K03031; AAA37332.1; JOINED.
DR   EMBL; K03032; AAA37332.1; JOINED.
DR   EMBL; K03033; AAA37332.1; JOINED.
DR   EMBL; K03034; AAA37332.1; JOINED.
DR   EMBL; K03035; AAA37332.1; JOINED.
DR   PIR; A23982; A23982.
DR   MGD; MGI:88467; Col1a1.
DR   InterPro; IPR000087; Collagen.
DR   InterPro; IPR000885; Fib_collagen_C.
DR   InterPro; IPR001007; VWF_C.
DR   Pfam; PF01391; Collagen; 18.
DR   Pfam; PF01410; COLFI; 1.
DR   ProDom; PD000007; Collagen; 1.
DR   ProDom; PD002078; Fib_collagen_C; 1.
DR   SMART; SM00038; COLFI; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Glycoprotein; Collagen; Signal.
FT   SIGNAL       1     22
FT   PROPEP      23    151       AMINO-TERMINAL PROPEPTIDE.
FT   CHAIN      152   1207       COLLAGEN ALPHA 1(I) CHAIN.
FT   PROPEP    1208   1453       CARBOXYL-TERMINAL PROPEPTIDE.
FT   DOMAIN      29     87       VWFC.
FT   DOMAIN     152    167       NONHELICAL REGION (N-TERMINAL).
FT   DOMAIN     168   1181       TRIPLE-HELICAL REGION.
FT   DOMAIN    1182   1207       NONHELICAL REGION (C-TERMINAL).
FT   CARBOHYD    56     56       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1354   1354       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   SITE       734    736       CELL ATTACHMENT SITE (POTENTIAL).
FT   SITE      1082   1084       CELL ATTACHMENT SITE (POTENTIAL).
FT   CONFLICT  1450   1450       A -> V (IN REF. 5).
SQ   SEQUENCE  1453 AA;  137944 MW;  3B802E535DF81808 CRC64;

  Query Match             6.3%;  Score 126;  DB 1;  Length 1453;
  Best Local Similarity   24.0%;  Pred. No. 8.7;
  Matches   66;  Conservative   30;  Mismatches   97;  Indels   82;  Gaps   12;

Qy        129 PPVFSRPISASYPPYQATGPPNTSYMPGMPGGISP----------YPSGYPPNPSGYPGC 178
              |   | |     ||  | | |    : |:|| | |          |:| || | | ||
Db       1120 PSGASGPAGPRGPPGSA-GSPGKDGLNGLPGPIGPPGPRGRTGDSGPAG-PPGPPGPPGP 1177

Qy        179 PYPPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDK-------LRWR 231
              | || | |  :   | | :        |:||   :   ||   : | |:       |:
Db       1178 PGPPSGGYDFSFLPQPPQE-------KSQDG---DRYYRADDANVVRDRDLAVDATLK-S 1226

Qy        232 MKEEMDRAQAELNALK---RTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEEL 288
              : :::: ::   : |   ||  |||  |   :     :|
Db       1227 LSQQIENIRSPEGSRKNPARTCRDLKMCHSDWKSGEYWID---------------------- 1266

Qy        289 SSALEKMENQSENNDIDEV----------IIPTAPLYKQILNLYAEENAIEDTIFYLGEA 338
                  ||  | |  :|          : || |   |  | |   |    : ||:
Db       1267 -------PNQGCNLDAIKVYCNMETGQTCVFPTQPSVPQ-KNWYISPNPKEKKHVWFGES 1318

Qy        339 LRRGV----------IDLDVFLKHVRLLSRKQFQ 362
              :  |            |: : |  :||:| :  |
Db       1319 MTDGFPFEYGSEGSDPTDVAIQLTFLRLMSTEASQ 1353


RESULT 13
YQY1_CAEEL
ID   YQY1_CAEEL     STANDARD;      PRT;   964 AA.
AC   Q09560;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Hypothetical 108.9 kDa protein F36G3.1 in chromosome X.
GN   F36G3.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   White S.;
RL   Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   REVISIONS.
RA   Jones S.J.M.;
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z47069; CAA87338.1; -.
DR   WormPep; F36G3.1; CE15979.
KW   Hypothetical protein; Coiled coil.
FT   DOMAIN      656    840       COILED COIL (POTENTIAL).
SQ   SEQUENCE   964 AA;  108868 MW;  DABA6BC59350F076 CRC64;

  Query Match             6.2%;  Score 124;  DB 1;  Length 964;
  Best Local Similarity   25.3%;  Pred. No. 7;
  Matches   50;  Conservative   46;  Mismatches   66;  Indels   36;  Gaps    9;
```

```
Qy        207 RDGTISEDTIRASLISAVSDKLRWRMKEEMDRAQAELNALKRTEEDLKKGHQKLEEMVTR 266
              :| |:  |     |:     ::   :: |:|| | :|  :|  |::| : : ||:| :
Db        636 QDQTVQTDNNSFILVDRHVPEVMESLQVEIDRLQGDLEKVKSGEKELLQINSKLKEELEE 695

Qy        267 LDQEV---------------AEVDKNIELLKKKDEELSSALEKMEN-----QSENNDID- 305
                | :               :|:|:  |::::||:||:   ||: |      |  :|:|
Db        696 SQQTIDGIEIEAEQQYTELTSEIDELCEIVQRKDQELAILKEKVTNVINIENSLKDDVDS 755

Qy        306 -EVIIPTAPLYKQIL-NLYAEENAIEDTIFYLGEA--LRRGVIDLDVFLKHVRLLSRKQF 361
              :||:      |:|: ||  | :||     |||   ||   :: :  |  :: : | :|
Db        756 QKVIVQR---QKEIIENLREELDAITKK---LGEVTKLRDKAVE-EATLYKMKNMERDRF 808

Qy        362 QLRALMQKARKTAGLSDL 379
                  | ::|: :  : ||
Db        809 ----LSREAQMSMEIEDL 822


RESULT 14
VASP_HUMAN
ID   VASP_HUMAN     STANDARD;      PRT;   380 AA.
AC   P50552; Q93035;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Vasodilator-stimulated phosphoprotein (VASP).
GN   VASP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=95129547; PubMed=7828592;
RA   Haffner C., Jarchau T., Reinhard M., Hoppe J., Lohmann S.M.,
RA   Walter U.;
RT   "Molecular cloning, structural analysis and functional expression of
RT   the proline-rich focal adhesion and microfilament-associated protein
RT   VASP.";
RL   EMBO J. 14:19-27(1995).
RN   [2]
RP   SEQUENCE OF 3-380 FROM N.A.
RX   MEDLINE=96411679; PubMed=8812448;
RA   Zimmer M., Fink T., Fischer L., Hauser W., Scherer K., Lichter P.,
RA   Walter U.;
RT   "Cloning of the VASP (vasodilator-stimulated phosphoprotein) genes in
RT   human and mouse: structure, sequence, and chromosomal localization.";
RL   Genomics 36:227-233(1996).
RN   [3]
RP   SEQUENCE OF 151-160; 235-244 AND 267-282, AND PHOSPHORYLATION SITES.
RX   MEDLINE=94237860; PubMed=8182057;
RA   Butt E., Abel K., Krieger M., Palm D., Hoppe V., Hoppe J., Walter U.;
RT   "cAMP- and cGMP-dependent protein kinase phosphorylation sites of the
RT   focal adhesion vasodilator-stimulated phosphoprotein (VASP) in vitro
RT   and in intact human platelets.";
RL   J. Biol. Chem. 269:14509-14517(1994).
RN   [4]
RP   BINDING TO PROFILIN.
RX   MEDLINE=95255215; PubMed=7737110;
RA   Reinhard M., Giehl K., Abel K., Haffner C., Jarchau T., Hoppe V.,
RA   Jockush B.M., Walter U.;
RT   "The proline-rich focal adhesion and microfilament protein VASP is a
RT   ligand for profilins.";
RL   EMBO J. 14:1583-1589(1995).
CC   -!- FUNCTION: ACTIN- AND PROFILIN-BINDING MICROFILAMENT-ASSOCIATED
CC       PROTEIN. MAY ACT IN CONCERT WITH PROFILIN TO CONVEY SIGNAL
CC       TRANSDUCTION TO ACTIN FILAMENT PRODUCTION.
CC   -!- SUBUNIT: HOMOTETRAMER.
CC   -!- SUBCELLULAR LOCATION: FOCAL ADHESIONS.
CC   -!- PTM: MAJOR SUBSTRATE FOR CAMP-DEPENDENT (CAPK) AND CGMP-DEPENDENT
CC       PROTEIN KINASE (CGPK) IN PLATELETS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z46389; CAA86523.1; -.
DR   EMBL; X98534; CAA67147.2; -.
DR   EMBL; X98533; CAA67147.2; JOINED.
DR   Genew; HGNC:12652; VASP.
DR   MIM; 601703; -.
DR   InterPro; IPR000697; RanBP1_WASP.
DR   InterPro; IPR001960; WH1.
DR   Pfam; PF00568; WH1; 1.
DR   SMART; SM00461; WH1; 1.
KW   Phosphorylation; Actin-binding.
FT   DOMAIN      118    122       POLY-PRO.
FT   DOMAIN      170    186       POLY-PRO.
FT   DOMAIN      215    222       POLY-GLY.
FT   DOMAIN      259    262       POLY-GLY.
FT   DOMAIN      322    325       POLY-SER.
FT   MOD_RES     157    157       PHOSPHORYLATION (BY PKA AND PKG).
FT   MOD_RES     239    239       PHOSPHORYLATION (BY PKA AND PKG).
FT   MOD_RES     278    278       PHOSPHORYLATION (BY PKA AND PKG).
SQ   SEQUENCE   380 AA;  39830 MW;  17634B8134DEBF59 CRC64;

  Query Match             6.2%;  Score 123.5;  DB 1;  Length 380;
  Best Local Similarity   19.6%;  Pred. No. 2.5;
  Matches   67;  Conservative   56;  Mismatches  112;  Indels  107;  Gaps   14;

Qy         33 VFNDGSSRELMNLTGTIPVPY-RGNTYNIPICLWLLDTYPYNPPICFVKPTSSMTIKTGK 91
              :::||: | |    || |  : |   |:                      ||::     |:
Db         15 LYDDGNKRWLP--AGTGPQAFSRVQIYH-------------------NPTANSFRVVGR 52

Qy         92 HVDANGKIYL-------------PYLHEWKHPQSDLLGLIQVMIVVFG--DEPPVFSRP 135
               :  : :: :              |  |:|:  :       ||  : ||  ::   |:
Db         53 KMQPDQQVVINCAIVRGVKYNQATPNFHQWRDAR-----QVWGLNFGSKEDAAQFAAG 105

Qy        136 ISASYPPYQATGPPNTSYMP--GMPGGISPY---------------------------- 164
              ::::    :  |||    :|   :| | ||
Db        106 MASALEALEGGGPPPPPALPTWSVPNGPSPEEVEQQKRQQPGPSEHIERRVSNAGGPPAP 165

Qy        165 PSGYPPNPSG---YPGCPYPPGGPYPATTSSQY------PSQPPVTTVGPSRDGTISEDT 215
              |:| || | |    || | ||| |     :: :     |  ||:           |
Db        166 PAGGPPPPPGPPPPPGPPPPPGLPPSGVPAAAHGAGGGPPPAPPLPAAQGPGGGGAGAPG 225

Qy        216 IRASLISAVSDKLRWRMKEE-------MDRAQA----------ELNA-LKRTEEDLKKGH 257
              : |::  |   |||   |:|        :|::          |:|| | |  :  : |
Db        226 LAAAIAGA---KLRKVSKQEEASGGPTAPKAESGRSGGGGLMEEMNAMLARRRKATQVGE 282

Qy        258 QKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 299
              :  ::     ::  | |    | :::  |: |: | :|:: |
Db        283 KTPKDESANQEEPEARVPAQSESVRRPWEKNSTTLPRMKSSS 324


RESULT 15
DIA1_HUMAN
ID   DIA1_HUMAN     STANDARD;      PRT;  1248 AA.
AC   O60610; Q9UC76;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Diaphanous protein homolog 1 (Diaphanous-related formin 1) (DRF1).
GN   DIAPH1 OR DIAP1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98028756; PubMed=9360932;
RA   Lynch E.D., Lee M.K., Morrow J.E., Welcsh P.L., Leon P.E., King M.-C.;
```

```
RT   "Nonsyndromic deafness DFNA1 associated with mutation of a human
RT   homolog of the Drosophila gene diaphanous.";
RL   Science 278:1315-1318(1997).
RN   [2]
RP   SEQUENCE OF 218-817 FROM N.A.
RC   TISSUE=Ovarian carcinoma;
RA   Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA   Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA   Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA   Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA   Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA   Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA   Ninomiya K., Iwayanagi T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 727-765 AND 1121-1145.
RC   TISSUE=Platelet;
RX   MEDLINE=95255215; PubMed=7737110;
RA   Reinhard M., Giehl K., Abel K., Haffner C., Jarchau T., Hoppe V.,
RA   Jockusch B.M., Walter U.;
RT   "The proline-rich focal adhesion and microfilament protein VASP is a
RT   ligand for profilins.";
RL   EMBO J. 14:1583-1589(1995).
CC   -!- FUNCTION: BINDS TO GTP-BOUND FORM OF RHO AND TO PROFILIN. ACTS IN
CC       A RHO-DEPENDENT MANNER TO RECRUIT PROFILIN TO THE MEMBRANE, WHERE
CC       IT PROMOTES ACTIN POLYMERIZATION. IT IS REQUIRED FOR CYTOKINESIS,
CC       STRESS FIBER FORMATION, AND TRANSCRIPTIONAL ACTIVATION OF THE
CC       SERUM RESPONSE FACTOR. DFR PROTEINS COUPLE RHO AND SRC TYROSINE
CC       KINASE DURING SIGNALING AND THE REGULATION OF ACTIN DYNAMICS (BY
CC       SIMILARITY). IN HEARING IT MAY PLAY A ROLE IN THE REGULATION OF
CC       ACTIN POLYMERIZATION IN HAIR CELLS.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, HEART, PLACENTA, LUNG,
CC       KIDNEY, PANCREAS, LIVER, SKELETAL MUSCLE, AND COCHLEA.
CC   -!- DOMAIN: DRFS ARE REGULATED BY INTRAMOLECULAR GBD-DAD BINDING WHERE
CC       RHO-GTP ACTIVATES THE DRFS BY DISRUPTING THE GBD-DAD INTERACTION
CC       (BY SIMILARITY).
CC   -!- DISEASE: DEFECTS IN DIAPH1 ARE A CAUSE OF AUTOSOMAL DOMINANT
CC       NONSYNDROMIC SENSORINEURAL DEAFNESS 1 (DFNA1).
CC   -!- SIMILARITY: CONTAINS 1 GTPASE-BINDING DOMAIN (GBD).
CC   -!- SIMILARITY: CONTAINS 1 FORMIN HOMOLOGY 1 (FH1) DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 FORMIN HOMOLOGY 2 (FH2) DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 FORMIN HOMOLOGY 3 (FH3) DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 DRF AUTOREGULATORY DOMAIN (DAD).
CC   -!- SIMILARITY: BELONGS TO THE FORMIN HOMOLOGY FAMILY. DIAPHANOUS
CC       SUBFAMILY.
CC   -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN THAT IT SEEMS
CC       TO INCLUDE INTRONIC SEQUENCE.
CC   -!- DATABASE: NAME=Hereditary hearing loss homepage;
CC       NOTE=Gene page;
CC       WWW="http://www.uia.ac.be/dnalab/hhh/hhhgenes.html".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF051782; AAC05373.1; -.
DR   EMBL; AK023345; BAB14533.1; ALT_SEQ.
DR   Genew; HGNC:2876; DIAPH1.
DR   MIM; 602121; -.
DR   MIM; 124900; -.
DR   InterPro; IPR003104; FH2.
DR   Pfam; PF02181; FH2; 1.
DR   SMART; SM00498; FH2; 1.
KW   Coiled coil; Repeat; Deafness.
FT   DOMAIN       63    260       GBD.
FT   DOMAIN      157    457       FH3.
FT   DOMAIN      460    563       COILED COIL (POTENTIAL).
FT   DOMAIN      588    743       FH1 (PRO-RICH).
FT   DOMAIN      748   1190       FH2.
FT   DOMAIN     1015   1172       COILED COIL (POTENTIAL).
FT   DOMAIN     1173   1187       DAD.
FT   DOMAIN     1189   1192       ARG/LYS-RICH (BASIC).
FT   CONFLICT    804    804       T -> TSKA (IN REF. 2).
FT   CONFLICT   1132   1133       RK -> AE (IN REF. 3).
SQ   SEQUENCE   1248 AA;  138978 MW;  ED1F5147CFF9A886 CRC64;

  Query Match             6.1%;  Score 123;  DB 1;  Length 1248;
  Best Local Similarity   22.0%;  Pred. No. 11;
  Matches   66;  Conservative   46;  Mismatches  124;  Indels   64;  Gaps   11;

Qy        129 PPVFSRPISASYP------PYQATGPPNTSYMPGMPG-----------GISPYPSGY--- 168
              ||    | ||  |      | :|  ||    :|| ||           || | | |
Db        668 PPPPPLPGSAGIPPPPPPLPGEAGMPPPPPPLPGGPGIPPPPPPFPGGPGIPPPPPGMGMP 727

Qy        169 PPNPSGYPGCPYPPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKL 228
              || | |: | |  |  |:  |   |  :| |    |:    ::||  :   : | :
Db        728 PPPPFGF-GVPAAPVLPFGLTPKKLY--KPEVQLRRPNWSKLVAEDLSQDCFWTKVKED- 783

Qy        229 RWRMKEEMDRAQAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEEL 288
              |:   |   :     :|  :|::| : | :|                | :: ||  | :
Db        784 RFENNELFAKLTLTFSAQTKTKKDQEGGEEKKSVQ----------KKKVKELKVLDSKT 832

Qy        289 SSALE------KMENQSENN---DIDEVIIPTAPLYKQILNLYAEE-----NAIEDTIFY 334
              :  |       :|  |   |   :::| ::  : :   |  :   |     : ::|
Db        833 AQNLSIFLGSFRMPYQEIKNVILEVNEAVLTESMIQNLIKQMPEPEQLKMLSELKDEYDD 892

Qy        335 LGEALRRGVIDLLDV--------------FLKHVRLLSRKQFQLRALMQKARKTAGLSDL 379
              | |: : ||:   |               | : |  :  :   : |  :: ||:   |:|
Db        893 LAESEQFGVVMGTVPRLRPRLNAILFKLQFSEQVENIKPEIVSVTAACEELRKSESFSNL 952


Search completed: June  3, 2003, 14:08:26
Job time : 15.9816 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: June 3, 2003, 14:06:00 ; Search time 52.9304 Seconds
(without alignments)
1479.263 Million cell updates/sec

Title: US-09-804-690-4
Perfect score: 2002
Sequence: 1 MVSKYKYRDLTVRETVNVIT.....FQLRALMQKARKTAGLSDLY 380

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1996	99.7	390	4	Q9BUM5	Q9bum5 homo sapien
2	1900.5	94.9	381	11	Q9CXS3	Q9cxs3 mus musculu
3	1862	93.0	392	13	Q9I8G8	Q9i8g8 chelonia my
4	1000.5	50.0	408	5	Q9GSB5	Q9gsb5 drosophila
5	809.5	40.4	331	5	Q9VVA7	Q9vva7 drosophila
6	658	32.9	404	5	O76258	O76258 caenorhabdi
7	455	22.7	398	10	Q9LHG8	Q9lhg8 arabidopsis
8	422	21.1	174	11	O08761	O08761 mus musculu
9	391	19.5	368	10	Q9FFY6	Q9ffy6 arabidopsis
10	371	18.5	177	4	Q96FF5	Q96ff5 homo sapien
11	371	18.5	341	4	Q9NUX7	Q9nux7 homo sapien
12	303	15.1	385	3	P87279	P87279 saccharomyc
13	169	8.4	376	3	P78998	P78998 saccharomyc
14	153.5	7.7	148	4	Q9NTQ8	Q9ntq8 homo sapien
15	148.5	7.4	148	4	Q9NZ81	Q9nz81 homo sapien
16	148.5	7.4	397	4	Q96DL9	Q96dl9 homo sapien
17	144.5	7.2	746	4	Q9BU60	Q9bu60 homo sapien
18	144.5	7.2	1052	4	Q96FT1	Q96ft1 homo sapien
19	142	7.1	1021	4	O15451	O15451 homo sapien
20	142	7.1	1251	4	O15450	O15450 homo sapien
21	141.5	7.1	503	11	Q921F1	Q921f1 mus musculu
22	141	7.0	247	10	Q945K9	Q945k9 arabidopsis
23	141	7.0	463	11	Q922A2	Q922a2 mus musculu
24	138.5	6.9	827	3	O60130	O60130 schizosacch
25	136	6.8	137	11	Q9CQJ5	Q9cqj5 mus musculu
26	135.5	6.8	437	10	Q9STN0	Q9stn0 arabidopsis
27	135.5	6.8	895	10	Q9LIW7	Q9liw7 oryza sativ
28	135	6.7	463	11	Q8VIN2	Q8vin2 rattus norv
29	135	6.7	674	5	Q18106	Q18106 caenorhabdi
30	134	6.7	201	16	O86316	O86316 mycobacteri
31	134	6.7	485	3	Q9UTP6	Q9utp6 schizosacch
32	132.5	6.6	471	10	Q9LD31	Q9ld31 crypthecodi
33	132.5	6.6	671	16	Q9Z502	Q9z502 streptomyce
34	131.5	6.6	370	3	P78821	P78821 schizosacch
35	131.5	6.6	900	10	Q9FIU0	Q9fiu0 arabidopsis
36	131.5	6.6	900	10	Q94B77	Q94b77 arabidopsis
37	131	6.5	472	3	O59907	O59907 neurospora
38	131	6.5	539	13	Q9I8F3	Q9i8f3 gallus gall
39	131	6.5	1880	5	O18465	O18465 hirudo medi
40	130.5	6.5	332	13	Q8QGD9	Q8qgd9 gallus gall
41	130.5	6.5	609	17	Q8TXA4	Q8txa4 methanopyru
42	130	6.5	345	11	Q9ESF4	Q9esf4 mus musculu
43	130	6.5	1647	4	Q9HBD3	Q9hbd3 homo sapien
44	130	6.5	1679	4	Q9HBD4	Q9hbd4 homo sapien
45	129.5	6.5	157	5	Q9W1K1	Q9w1k1 drosophila

ALIGNMENTS

RESULT 1
Q9BUM5
ID Q9BUM5 PRELIMINARY; PRT; 390 AA.
AC Q9BUM5;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Tumor susceptibility gene 101.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EYE;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; BC002487; AAH02487.1; -.
DR InterPro; IPR000608; UBQ_conjugat..
DR SMART; SM00212; UBCc; 1.
SQ SEQUENCE 390 AA; 43910 MW; 96BB2A4FC22DF16A CRC64;

Query Match 99.7%; Score 1996; DB 4; Length 390;
Best Local Similarity 99.7%; Pred. No. 9e-120;
Matches 379; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          11 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 70

Qy          61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          71 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 130

Qy         121 MIVVFGDEPPVFSRPISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCPY 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         131 MIVVFGDEPPVFSRPISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCPY 190

Qy         181 PPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRAQ 240
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        191 PPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRAQ 250

Qy        241 AELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQSE 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        251 AELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQSE 310

Qy        301 NNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRKQ 360
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        311 NNDIDEVIIPTAPLYKQILNLYAEENAIEDTILYLGEALRRGVIDLDVFLKHVRLLSRKQ 370

Qy        361 FQLRALMQKARKTAGLSDLY 380
              ||||||||||||||||||||
Db        371 FQLRALMQKARKTAGLSDLY 390


RESULT 2
Q9CXS3
ID   Q9CXS3      PRELIMINARY;      PRT;   381 AA.
AC   Q9CXS3;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Tumor susceptibility gene 101.
GN   TSG101.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=EMBRYONIC HEAD;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK014049; BAB29131.1; -.
DR   MGD; MGI:106581; tsg101.
DR   InterPro; IPR000608; UBQ_conjugat.
DR   SMART; SM00212; UBCc; 1.
SQ   SEQUENCE   381 AA;  43021 MW;  FA650A0BF7B21ABF CRC64;

  Query Match           94.9%;  Score 1900.5;  DB 11;  Length 381;
  Best Local Similarity   94.5%;  Pred. No. 1.1e-113;
  Matches  360;  Conservative   11;  Mismatches    9;  Indels    1;  Gaps    1;

Qy          1 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 60
              |:|||||||||||:|||| :|||||||||||||||||||:|||||||| |||| |||
Db          1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNI 60

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120
              ||||||||||||||||||||||||||||||||||||||||||||:||||:|:|| |||:
Db         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQI 120

Qy        121 MIVVFGDEPPVFSRP-ISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCP 179
              |||:||:|||||||| :|||||||-||||||||||||| ||| ||||||||||||||||
```

```
Db        121 MIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCP 180

Qy        180 YPPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRA 239
              ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        181 YPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGA 240

Qy        240 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 300

Qy        300 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 360

Qy        360 QFQLRALMQKARKTAGLSDLY 380
              ||||||||||||||||||||
Db        361 QFQLRALMQKARKTAGLSDLY 381


RESULT 3
Q9I8G8
ID   Q9I8G8      PRELIMINARY;      PRT;   392 AA.
AC   Q9I8G8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Tumor susceptibility protein 101.
GN   TSG101.
OS   Chelonia mydas caranigra (Green sea-turtle).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Testudines; Cryptodira; Chelonioidea; Cheloniidae; Chelonia.
OX   NCBI_TaxID=8469;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Yu Q., Lu Y., Nerurkar V.R., Yanagihara R.;
RT   "Studies on the turtle tumor susceptibility gene, TSG; full-length
RT   cDNA sequence, genomic structure analysis and role in green turtle
RT   fibropapilloma.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF279276; AAF87776.1; -.
DR   InterPro; IPR000608; UBQ_conjugat.
DR   SMART; SM00212; UBCc; 1.
SQ   SEQUENCE   392 AA;  44314 MW;  25DE98A5116CF8EA CRC64;

  Query Match           93.0%;  Score 1862;  DB 13;  Length 392;
  Best Local Similarity   92.7%;  Pred. No. 3.2e-111;
  Matches  354;  Conservative   15;  Mismatches   11;  Indels    2;  Gaps    2;

Qy          1 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 60
              |::|||||||||:|| :||| |||||||:|:||||||||:||:|||||||||||||||||
Db         11 MLAKYKYRDLTVQETTSVITQYKDLKPVMDAYVFNDGSSRDLMSLTGTIPVPYRGNTYNI 70

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120
              ||||||||||:|||||||||||||||||||||||||||||||||||||||||||:||||:
Db         71 PICLWLLDTYPFNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLIGLIQI 130

Qy        121 MIVVFGDEPPVFSRP-ISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCP 179
              ||||||:|||||||| || |: ||||||||||||||||| |||||| |:|||||||| |
Db        131 MIVVFGEEPPVFSRPTISTSFQPYQATGPPNTSYMPGMPSGISPYPPGHPPNPSGYPGYP 190

Qy        180 YPPGGPYPATTSSQ-YPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDR 238
              ||||||:|||||  | | ||||||||||||||||||||| ||||||||||||||||||
Db        191 YPPGGPFPATTSGQHYTSQPPVTTVGPSRDGTISEDTIRASRISAVSDKLRWRMKEEMDR 250

Qy        239 AQAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQ 298
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        251 AQAELNALKRTEEDLKKGHQKLEEMVTRLDHEVAEVDKNIELLKKKDEELSSALEKMENQ 310

Qy        299 SENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSR 358
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        311 SENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSR 370

Qy        359 KQFQLRALMQKARKTAGLSDLY 380
```

```
                    ||||||||||||||||||||||
Db        371 KQFQLRALMQKARKTAGLSDLY 392


RESULT 4
Q9GSB5
ID   Q9GSB5      PRELIMINARY;      PRT;   408 AA.
AC   Q9GSB5;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Tumor suppressor protein 101.
GN   TSG101 OR CG9712.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Laurencon A., Burtis K.C., Hawley S.;
RT   "TSG101 fly homolog.";
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF315343; AAG29564.1; -.
DR   FlyBase; FBgn0036666; TSG101.
DR   InterPro; IPR001440; TPR.
DR   InterPro; IPR000608; UBQ_conjugat.
DR   SMART; SM00212; UBCc; 1.
SQ   SEQUENCE   408 AA;  45190 MW;  7C48E20DEBB29A19 CRC64;

  Query Match             50.0%;  Score 1000.5;  DB 5;  Length 408;
  Best Local Similarity   50.6%;  Pred. No. 3.3e-56;
  Matches  203;  Conservative   63;  Mismatches  104;  Indels   31;  Gaps    7;

Qy          2 VSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNIP 61
              :|||||   | :: |:|:| :: |   |  :|||||||:||  : |||||  |: ||| ||
Db         13 LSKYKYVAATKKDVVDVVTSFRSLTYDLQRFVFNDGSSKELFTIQGTIPVVYKNNTYYIP 72

Qy         62 ICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQVM 121
              ||:||:||:| | |:|||||| :| ||   :|| |||:||||||:|:   |||| |||||
Db         73 ICIWLMDTHPQNAPMCFVKPTPTMQIKVSMYVDHNGKVYLPYLHDWQPHSSDLLSLIQVM 132

Qy        122 IVVFGDEPPVFSRPISASYPPYQATGPPNTSYM--PGMPGGIS---PYP-------SGYP 169
              || ||| |||:|:|      ||     |  |||  || ||| :   |||       | :|
Db        133 IVTFGDHPPVYSKPKEQIAAPY-----PTNSYMPQPGAPGGSNSFLPYPTAGGAGGSNFP 187

Qy        170 PNPSGYPGCPYPP-------GGPYPA------TTSSQYPSQPPVTTVGPSRDGTISEDTI 216
              | |:|    ||||       |  |||       |:  ||          ||  |||:|: |
Db        188 PYPTGSNVGPYPPTPAGPAGGSGYPAYPNFIQPTAGGYPPAAGYNPSNPSSTGTITEEHI 247

Qy        217 RASLISAVSDKLRWRMKEEMDRAQAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDK 276
              :||:|||: |||| |::|:::: |||:  | ||:::| :|  |:: :: ||::|  :: |
Db        248 KASIISAIDDKLRRRVQEKVNQYQAEIETLNRTKQELLEGSAKIDAIIERLEREHIDMQK 307

Qy        277 NIELLKKKDEELSSALEKMENQSENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLG 336
              || :|| |::||  ||| :|:    |  || :  |||||:|:|| ||:| | || |:|||
Db        308 NISILKDKEQELEKALEDLESAEAINP-DEAVTTTAPLYRQLLNAYADEAATEDAIYYLG 366

Qy        337 EALRRGVIDLDVFLKHVRLLSRKQFQLRALMQKARKTAGLS 377
              | || |||||: |||||| |||||| ||| ||| |: |||:
Db        367 EGLRGGVIDLETFLKHVRQLSRKQFILRATMQKCRQKAGLA 407


RESULT 5
Q9VVA7
ID   Q9VVA7      PRELIMINARY;      . PRT;   331 AA.
AC   Q9VVA7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CG9712 protein.
GN   TSG101 OR CG9712.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003526; AAF49406.1; -.
DR   FlyBase; FBgn0036666; TSG101.
DR   InterPro; IPR001440; TPR.
SQ   SEQUENCE   331 AA;  36255 MW;  A6DDDA84C718B738 CRC64;

  Query Match             40.4%;  Score 809.5;  DB 5;  Length 331;
  Best Local Similarity   50.0%;  Pred. No. 4e-44;
  Matches  168;  Conservative   51;  Mismatches   86;  Indels   31;  Gaps    7;

Qy         67 LDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQVMIVVFG 126
              :||:| | |:|||||| :| ||   :|| |||:|||||:|:   |||| ||||||| ||
Db          1 MDTHPQNAPMCFVKPTPTMQIKVSMYVDHNGKVYLPYLHDWQPHSSDLLSLIQVMIVTFG 60

Qy        127 DEPPVFSRPISASYPPYQATGPPNTSYM--PGMPGGIS---PYP-------SGYPPNPSG 174
              | |||:|:|      ||     |  |||  || ||| :   |||       | :|| |:|
Db         61 DHPPVYSKPKEQIAAPY-----PTNSYMPQPGAPGGSNSFLPYPTAGGAGGSNFPPYPTG 115

Qy        175 YPGCPYPP-------GGPYPA------TTSSQYPSQPPVTTVGPSRDGTISEDTIRASLI 221
                 ||||       |  |||       |:  ||          ||  |||:|: |:||:|
Db        116 SNVGPYPPTPAGPAGGSGYPAYPNFIQPTAGGYPPAAGYNPSNPSSTGTITEEHIKASII 175

Qy        222 SAVSDKLRWRMKEEMDRAQAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELL 281
              ||: |||| |::|:::: |||:  | ||:::| :|  |:: :: ||::|  :: ||| :|
Db        176 SAIDDKLRRRVQEKVNQYQAEIETLNRTKQELLEGSAKIDAIIERLEREHIDMQKNISIL 235

Qy        282 KKKDEELSSALEKMENQSENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRR 341
              | |::||  ||| :|:    |  || :  |||||:|:|| ||:| | || |:|||| ||
```

```
Db        236 KDKEQELEKALEDLESAEAINP-DEAVTTTAPLYRQLLNAYADEAATEDAIYYLGEGLRG 294

Qy        342 GVIDLDVFLKHVRLLSRKQFQLRALMQKARKTAGLS 377
              |||||: |||||| |||||| ||| ||| |: |||:
Db        295 GVIDLETFLKHVRQLSRKQFILRATMQKCRQKAGLA 330


RESULT 6
O76258
ID   O76258      PRELIMINARY;      PRT;   404 AA.
AC   O76258;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   C09G12.9 protein.
GN   C09G12.9.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Le T.T., Harper M.;
RT   "The sequence of C. elegans cosmid C09G12.";
RL   Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF038608; AAC25822.1; -.
DR   InterPro; IPR000608; UBQ_conjugat.
DR   SMART; SM00212; UBCc; 1.
SQ   SEQUENCE   404 AA;  43909 MW;  2C322EC52DF4C62D CRC64;

  Query Match           32.9%;  Score 658;  DB 5;  Length 404;
  Best Local Similarity   36.7%;  Pred. No. 2.3e-34;
  Matches  151;  Conservative   74;  Mismatches  121;  Indels   66;  Gaps   11;

Qy          6 KYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNIPICLW 65
              || |   :: :  :: :||| |  |:::| ||  |    | ||||| |:|  ||||: ::
Db         16 KYADSAKKDIIGALSQFKDLSPGTDTFMFPDGKRRTAFRLKGTIPVYYKGACYNIPVTVY 75

Qy         66 LLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQVMIVVF 125
              | ||:||  |||:| |||:|     :||:  ||::||||:||: |  || ||:|:
Db         76 LWDTHPYYAPICYVNPTSTM---ESEHVNKEGKVFLPYLNEWRFPGYDLSGLLQM----- 127

Qy        126 GDEPPVFSRPISASYPPYQATGP----------PNTSYMPGMPGGISPYPSG-----YP 169
                 :|:|  : |     || |           |  |  | ||   :|||:|     ||
Db        128 -----IFARSAANSATNASATNPSAGSSASSTPTPYPSSQPTMP---TPYPTGSGAAPYP 179

Qy        170 PNPSGYPGC--------------PYPPG----GPYPATTSSQY---PSQPPV---TTVG 204
              |: : ||                |||  |     |||: :|:     |  |||   |:|
Db        180 PSSTPYPSAGAMGYNPYMNVPQSTPYPMGASGASPYPSASSNPAPPPPRPPPVTAQTSVS 239

Qy        205 PSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRAQAELNALKRTEEDLKKGHQKLEEMV 264
               |  |||  |||||||::||| :|:| :::| |    ||: ::: | ::|::| |||: |:
Db        240 SSSGGTIQADTIRASVMSAVEEKIRAKLRERMGTNSAEMASIRTTSDELREGQQKLKRML 299

Qy        265 TRLDQEVAEVDKNIELLKKKDEELSSALEKMENQSENNDIDEVIIPTAPLYKQILNLYAE 324
              |: : : :    |:   |  ||: ||      ::   ||| |    ||::||:  ||:
Db        300 EELETQRSSLQTACEIYTAKKAELAKALSD-AGGTDAPPIDEAIDAAFPLHRQIVLNYAK 358

Qy        325 ENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRKQFQLRALMQKARKTAGL 376
```

```
                :          ::|::  | |  :|:||| :||:|| || ||| |:||||
Db        359 DLTC--------QSLKKRQITLAEYLRHVRDVSREQFIYRATMQKCRRTAGL 402


RESULT 7
Q9LHG8
ID   Q9LHG8      PRELIMINARY;      PRT;   398 AA.
AC   Q9LHG8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Mouse and human tumor susceptibility gene-like protein (Hypothetical
DE   44.7 kDa protein).
GN   T2E22.28 OR T2E22_129/AT3G12400.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RA   Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RX   PubMed=10907853;
RA   Nakamura Y.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT   Sequence features of the regions of 4,251,695 bp covered by ninety P1,
RT   TAC and BAC clones.";
RL   DNA Res. 7:217-221(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE=21016720; PubMed=11130713;
RA   Salanoubat M., Lemcke K., Rieger M., Ansorge W., Unseld M.,
RA   Fartmann B., Valle G., Bloecker H., Perez-Alonso M., Obermaier B.,
RA   Delseny M., Boutry M., Grivell L.A., Mache R., Puigdomenech P.,
RA   De Simone V., Choisne N., Artiguenave F., Robert C., Brottier P.,
RA   Wincker P., Cattolico L., Weissenbach J., Saurin W., Quetier F.,
RA   Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
RA   Wurmbach E., Drzonek H., Erfle H., Jordan N., Bangert S.,
RA   Wiedelmann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
RA   Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simionati B.,
RA   Conrad A., Hornischer K., Kauer G., Loehnert T.-H., Nordsiek G.,
RA   Reichelt J., Scharfe M., Schoen O., Bargues M., Terol J., Climent J.,
RA   Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,
RA   Cooke R., Laudie M., Berger-Llauro C., Purnelle B., Masuy D.,
RA   de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
RA   Monfort A., Argiriou A., Flores M., Liguori R., Vitale D.,
RA   Mannhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,
RA   Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
RA   Rooney T., Rizzo M., Walts A., Utterback T., Fujii C.Y., Shea T.P.,
RA   Creasy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
RA   Pai G., Militscher J., Sellers P., Gill J.E., Feldblyum T.V.,
RA   Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
RA   Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
RA   Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
RA   Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA   Nakayama S., Nakazaki N., Shinpo S., Takeuchi C., Wada T.,
RA   Watanabe A., Yamada M., Yasuda M., Tabata S.;
RT   "Sequence and analysis of chromosome 3 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 408:820-822(2000).
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Yamada K., Banh J., Banno F., Dale J.M., Goldsmith A.D., Lee J.M.,
RA   Onodera C.S., Quach H.L., Tang C., Toriumi M., Yamamura Y., Yu G.,
RA   Yu S., Bowser L., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,
RA   Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Kim C.,
RA   Koesema E., Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M.,
```

```
RA   Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shinn P.,
RA   Southwick A., Tracy S.E., Shinozaki K., Davis R.W., Ecker J.R.,
RA   Theologis A.;
RT   "Full Length cDNA of gene T2E22_129/At3g12400 (GI:15230478).";
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP002047; BAB03147.1; -.
DR   EMBL; AC069474; AAG51025.1; -.
DR   EMBL; AY056283; AAL07132.1; -.
DR   InterPro; IPR000608; UBQ_conjugat.
DR   SMART; SM00212; UBCc; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   398 AA;  44715 MW;  659996F9784FDC20 CRC64;

  Query Match             22.7%;  Score 455;  DB 10;  Length 398;
  Best Local Similarity   32.1%;  Pred. No. 2.1e-21;
  Matches  117;  Conservative   69;  Mismatches  146;  Indels   32;  Gaps   10;

Qy         16 VNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNIPICLWLLDTYPYNPP 75
              :|:|: |  |:|   |:: ||| |  |:   ||||:|: | |||||: :|||::|| :||
Db         42 LNLISSYPSLEPKTASFMHNDGRSVNLLQADGTIPMPFHGVTYNIPVIIWLLESYPRHPP 101

Qy         76 ICFVKPTSSMTIK-TGKHVDANGKIYLPYLHEWKHPQSDLLGLIQVMIVVFGDEPPVFSR 134
                :| ||: | ||    ||  :| : |||| | :| |:|: |:  :   |  :||::||
Db        102 CVYVNPTADMIIKRPHAHVTPSGLVSLPYLQNWVYPSSNLVDLVSDLSAAFARDPPLYSR 161

Qy        135 PISASYPPYQATGPPNT--SYMPGMPGGISPYPSGYPPNPSGYPGCPYPPGGPYPATTSS 192
                   ||     ||    | :   |       |  :||:| |          ||
Db        162 RRPQPPPP----SPPTVYDSSLSRPPSADQSLPRPFPPSPYG--------GGVSRVQVQH 209

Qy        193 QYPSQPPVTTVGPSRDGTISE--DTIRASLISAVSDKLRWRMKEEMDRAQAELNALKRTE 250
              :  |         :   |::  : : : |:|    : |    ||:   ||    ||| |
Db        210 VHHQQQSDDAAEVFKRNAINKMVEMVHSDLVS--MRRAREAEAEELLSLQA---GLKRRE 264

Qy        251 EDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQSEN---NDIDEV 307
              ::|  |   |:|||    :|   ::: ::::    : | | : : : :::|    |:|
Db        265 DELNIG---LKEMV----EEKETLEQQLQIISMNTDILDSWVRENQGKTKNLVDLDVDNA 317

Qy        308 IIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLLDVFLKHVRLLSRKQFQLRALM 367
                  | ||:|   | : |||| |: | :: : ||:  | :|::|||||:|| ||
Db        318 FECGDTLSKQMLECTALDLAIEDAIYSLDKSFQDGVVPFDQYLRNVRLLSREQFFHRATG 377

Qy        368 QKAR 371
               | |
Db        378 SKVR 381


RESULT 8
O08761
ID   O08761      PRELIMINARY;      PRT;   174 AA.
AC   O08761;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Signaling molecule.
GN   ATTP OR ATTP.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=THYMUS;
RA   Burbelo P.D.;
RT   "Role of ATTP in signal transduction.";
RL   Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U97571; AAB57699.1; -.
DR   MGD; MGI:1860490; Attp.
DR   InterPro; IPR000608; UBQ_conjugat.
DR   SMART; SM00212; UBCc; 1.
SQ   SEQUENCE   174 AA;  20499 MW;  81BEE363DD81900B CRC64;

  Query Match             21.1%;  Score 422;  DB 11;  Length 174;
  Best Local Similarity   54.8%;  Pred. No. 9.9e-20;
  Matches   74;  Conservative   26;  Mismatches   35;  Indels    0;  Gaps    0;

Qy          1 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 60
              :: |||:|||| |  ||   :   :  :|:||| | | ::|:| |||||| |:| ||||
Db         11 LLGKYKFRDLTVEELKNVSVSFPHFRYSVDTYVFKDTSQKDLLNFTGTIPVMYQGKTYNI 70

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120
              ||  |:||::|: |||||:|||::| |  |||||| |:|||||  | ||:| ::|||:
Db         71 PIRFWILDSHPFAPPICFLKPTANMEISVGKHVDAKGRIYLPYLQNWSHPKSAIVGLIKE 130

Qy        121 MIVVFGDEPPVFSRP 135
              ||  | :| |::| |
Db        131 MIARFQEELPLYSIP 145


RESULT 9
Q9FFY6
ID   Q9FFY6      PRELIMINARY;      PRT;   368 AA.
AC   Q9FFY6;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Similarity to nascent polypeptide associated complex alpha chain.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RX   MEDLINE=97471969; PubMed=9330910;
RA   Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,
RA   Miyajima N., Tabata S.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence
RT   features of the 1.6 Mb regions covered by twenty physically assigned
RT   P1 clones.";
RL   DNA Res. 4:215-230(1997).
DR   EMBL; AB005230; BAB11114.1; -.
DR   InterPro; IPR000608; UBQ_conjugat.
DR   SMART; SM00212; UBCc; 1.
SQ   SEQUENCE   368 AA;  41495 MW;  262F8CDB45BBCD40 CRC64;

  Query Match             19.5%;  Score 391;  DB 10;  Length 368;
  Best Local Similarity   28.4%;  Pred. No. 2.3e-17;
  Matches  108;  Conservative   67;  Mismatches  129;  Indels   76;  Gaps   11;

Qy          8 RDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNIPICLWLL 67
              : |  :  :|:|: |  | |   ::  ||| |  |:   ||||:|:::| :|||: :|||
Db         34 KSLIRQHLLNLISSYTSLDPKTATFTHNDGRSVILLQADGTIPMPFQGVSYNIPVVIWLL 93

Qy         68 DTYPYNPPICFVKPTSSMTIK-TGKHVDANGKIYLPYLHEWKHPQSDLLGLIQVMIVVFG 126
              ::||  ||  :| ||  | ||     :|  :| : ||||  | :| |:|: |   :   |
Db         94 ESYPQYPPCVYVNPTRDMIIKRPHSNVSPSGLVSLPYLQNWIYPSSNLVDLASHLSAAFS 153

Qy        127 DEPPVFS--RPISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCPYP-PG 183
              :||::|  ||                                  || ||   | |  |
Db        154 RDPPLYSQRRP--------------------------------PPQPSPSIGSGYSRPL 180

Qy        184 GPYPATTSSQYPSQPPVTTVGPSRDGTI-----SEDTIRASLISAVSDKLRWRMKEEMDR 238
              |     :::   :  :  :     | |     : :     |:|  ||
Db        181 PPRQTDDAAEVYKKNAINRIVEMVHGDIVLMRSAREVETEGLLSLQSD------------ 228

Qy        239 AQAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQ 298
                  ||| ||::  |    :|||  :::|  |  : ::::    : | | :   |||
Db        229 -------LKRREEEINNG---FKEMV--IEKETLE--QQLQVIAMNTDVLGSWIR--ENQ 272

Qy        299 SE------NNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLLDVFLKH 352
              :       : |:|:       | ||:|   | : |||| :: : :: | | :  | :|::
Db        273 GKAKDLLVDLDVDDSFECIDSLSKQMLECTALDLAIEDVVYSMDKSFRDGSLPFDQYLRN 332

Qy        353 VRLLSRKQFQLRALMQKARK 372
              ||||||:||  ||  :| |:
Db        333 ------------------ 332
```

```
Db        333 VRLLSREQFFHRATAEKVRE 352


RESULT 10
Q96FF5
ID   Q96FF5      PRELIMINARY;      PRT;   177 AA.
AC   Q96FF5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Unknown (protein for MGC:13394).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RA   Strausberg R.;
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC011011; AAH11011.1; -.
SQ   SEQUENCE   177 AA;  19552 MW;  E733E710F8D0068A CRC64;

  Query Match             18.5%;  Score 371;  DB 4;  Length 177;
  Best Local Similarity   60.0%;  Pred. No. 1.8e-16;
  Matches   63;  Conservative   21;  Mismatches   21;  Indels    0;  Gaps    0;

Qy         29 LDSYVFNDGSSRELMNLTGTIPVPYRGNTYNIPICLWLLDTYPYNPPICFVKPTSSMTIK 88
              :|:||| | | ::|:| |||||| |:|||||||  |:||::|: |||||:|||::| |
Db          1 MDTYVFKDSSQKDLLNFTGTIPVMYQGNTYNIPIRFWILDSHPFAPPICFLKPTANMGIL 60

Qy         89 TGKHVDANGKIYLPYLHEWKHPQSDLLGLIQVMIVVFGDEPPVFS 133
              |||||| |:|||||  | ||:| ::|||: ||  | :| |::|
Db         61 VGKHVDAQGRIYLPYLQNWSHPKSVIVGLIKEMIAKFQEELPMYS 105


RESULT 11
Q9NUX7
ID   Q9NUX7      PRELIMINARY;      PRT;   341 AA.
AC   Q9NUX7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   CDNA FLJ11068 fis, clone PLACE1004918, weakly similar to L-lactate
DE   dehydrogenase M chain (EC 1.1.1.27).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RA   Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA   Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA   Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA   Nakamura Y., Nagahari K., Masuho Y., Sasaki N.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK001930; BAA91985.1; -.
DR   HSSP; P00336; 5LDH.
DR   InterPro; IPR001236; ldh.
DR   InterPro; IPR001557; L_LDH.
DR   InterPro; IPR000205; NAD_binding.
DR   InterPro; IPR000608; UBQ_conjugat.
DR   Pfam; PF00056; ldh; 1.
DR   Pfam; PF02866; ldh_C; 1.
DR   PRINTS; PR00086; LLDHDRGNASE.
DR   SMART; SM00212; UBCc; 1.
SQ   SEQUENCE   341 AA;  37587 MW;  7D5CD535296165FA CRC64;

  Query Match             18.5%;  Score 371;  DB 4;  Length 341;
  Best Local Similarity   60.0%;  Pred. No. 4e-16;
  Matches   63;  Conservative   21;  Mismatches   21;  Indels    0;  Gaps    0;

Qy         29 LDSYVFNDGSSRELMNLTGTIPVPYRGNTYNIPICLWLLDTYPYNPPICFVKPTSSMTIK 88
              :|:||| | | ::|:| |||||| |:|||||||  |:||::|: |||||:|||::| |
Db          1 MDTYVFKDSSQKDLLNFTGTIPVMYQGNTYNIPIRFWILDSHPFAPPICFLKPTANMGIL 60

Qy         89 TGKHVDANGKIYLPYLHEWKHPQSDLLGLIQVMIVVFGDEPPVFS 133
              |||||| |:|||||  | ||:| ::|||: ||  | :| |::|
Db         61 VGKHVDAQGRIYLPYLQNWSHPKSVIVGLIKEMIAKFQEELPMYS 105


RESULT 12
P87279
ID   P87279      PRELIMINARY;      PRT;   385 AA.
AC   P87279;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   STP22P.
GN   STP22 OR YCL008C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=XG1#1;
RA   Li Y., Kane T., Tipper C., Spatrick P., Jenness D.D.;
RT   "The complete sequence of STP22 gene.";
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF004731; AAB62820.1; -.
DR   SGD; S0000514; STP22.
DR   InterPro; IPR000608; UBQ_conjugat.
DR   SMART; SM00212; UBCc; 1.
DR   PROSITE; PS50127; UBIQUITIN_CONJUGAT_2; 1.
SQ   SEQUENCE   385 AA;  43330 MW;  FEDE3BE79F063BCE CRC64;

  Query Match             15.1%;  Score 303;  DB 3;  Length 385;
  Best Local Similarity   26.2%;  Pred. No. 1e-11;
  Matches  102;  Conservative   77;  Mismatches  163;  Indels   48;  Gaps   15;

Qy          7 YRD--LTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNT-YNIPIC 63
              | |   |  ::: ::  :  |:|    :  :||: : |::: |||     |:: ::||:
Db         25 YNDGRTTFHDSLALLDNFHSLRPRTRVFTHSDGTPQLLLSIYGTISTGEDGSSPHSIPVI 84

Qy         64 LWLLDTYPYNPPIC------FVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGL 117
              :|:   ||  ||       |   | | ::   :::|:|| | || || |     :|: :
Db         85 MWVPSMYPVKPPFISINLENFDMNTISSSLPIQEYIDSNGWIALPILHCWDPAAMNLIMV 144

Qy        118 IQVMIVVFGDEPPVFSRPISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPG 177
              :| :: :   |||    |     |  |     ||  :|  |      |   || |
Db        145 VQELMSLL-HEPPQDQAPSLPPKPNTQLQQEQNTPPLPPKPKS----PHLKPPLP----- 194

Qy        178 CPYPPGGPYPATTSSQYPSQPPVTTVGPSRD-------GTISEDTIRASLISAVSDKLRW 230
              ||  | ||: :          | : |:          |:  :  |   :  |:||:
Db        195 ---PPPPPQPASNALDLMDMDN-TDISPTNHHEMLQNLQTVVNELYRED-VDYVADKILT 249

Qy        231 R---MKEEMDRAQAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEE 287
              |   |:| : |   |: |: :    |:   | :|: :  |: ::  :  |    : | ::
Db        250 RQTVMQESIARFH-EIIAIDKNH--LRAVEQAIEQTMHSLNAQIDVLTAN----RAKVQQ 302

Qy        288 LSSALEKMENQSENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLD 347
              ||      :  :: |:: : :          |: || |:: |: |||  |   | || | ||
Db        303 FSST-----SHVDDEDVNSIAVAKTDGLNQLYNLVAQDYALTDTIECLSRMLHRGTIPLD 357

Qy        348 VFLKHVRLLSRKQFQLRALMQKARKTAGLS 377
              |:|  | |:|:|| :|  :|  | |: ||
Db        358 TFVKQGRELARQQFLVRWHIQ--RITSPLS 385


RESULT 13
P78998
ID   P78998      PRELIMINARY;      PRT;   376 AA.
AC   P78998;
```

```
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Hypothetical 41.5 kDa protein.
GN   LGYCLUN5C.
OS   Saccharomyces pastorianus (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=27292;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Andersen T., Nilsson-Tillgren T.;
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z86109; CAB06793.1; -.
DR   InterPro; IPR001284; Ribosomal_L34E.
DR   InterPro; IPR000608; UBQ_conjugat.
DR   SMART; SM00212; UBCc; 1.
DR   PROSITE; PS01145; RIBOSOMAL_L34E; UNKNOWN_1.
DR   PROSITE; PS50127; UBIQUITIN_CONJUGAT_2; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   376 AA;  41542 MW;  837F91DA97BF0A0B CRC64;

  Query Match             8.4%;  Score 169;  DB 3;  Length 376;
  Best Local Similarity   19.0%;  Pred. No. 0.0035;
  Matches   60;  Conservative   65;  Mismatches  140;  Indels   50;  Gaps    9;

Qy         11 TVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNIPICLWLLDTY 70
              |  ::: ::  :  |:|    :  :||| : |::: ||:     |:  ::|: :|:   |
Db         31 TFHDSLALLDKFHRLRPRTRVFTHSDGSPQLLLSIYGTV-----GD--SLPLIMWIPSLY 83

Qy         71 PYNPPI------CFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQVMIVV 124
              |  ||        |     | ::    ::|::|:: || |  |      |: ::| :: :
Db         84 PVKPPFISIDLETFDVSAISSSLPVQAYIDSDGRVALPILDHWNPATMGLIVVVQELMSL 143

Qy        125 FGDEPPVFSRPISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCPYPPGG 184
              :                     |:   :| :|    | |     :|  :   | ||
Db        144 LSE--------------------PSRDRVPSLP----PKPDAM-HSPLKHTVPPLPPKA 177

Qy        185 PYPATTSSQYPSQPPVTTVGPS-------RDGTISEDTIRASLISAVSDKLRWRMKEEMD 237
               :|     | | |||     ||        | |    |    ::  :   |    :|::
Db        178 KFPHV---QPPLQPPPPPQPPSSAVDLMDMDNTDLSPTNHHEMLQNLQSVLNELYREDVH 234

Qy        238 R-AQAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKME 296
              | |      ::  : : |: :    |||      ::: :  |  : | |::    ::
Db        235 YVADKILTRQTIMQDSVARFHEMVAVDKTRLQAVEQTIEQTMHTLNAQIEVLTAERATVQ 294

Qy        297 NQSENNDIDEVIIPT 311
              |  :  |:  : |
Db        295 EFSSTSPFDDEDVDT 309


RESULT 14
Q9NTQ8
ID   Q9NTQ8      PRELIMINARY;      PRT;   148 AA.
AC   Q9NTQ8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Hypothetical 15.4 kDa protein.
GN   DKFZP564J157.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=21154917; PubMed=11230166;
RA   Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,
RA   Ansorge W., Boecher M., Bloecker H., Bauersachs S., Blum H.,
RA   Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,
RA   Mewes H.W., Ottenwaelder B., Obermaier B., Tampe J., Heubner D.,
RA   Wambutt R., Korn B., Klein M., Poustka A.;
RT   "Towards a Catalog of Human Genes and Proteins: Sequencing and
RT   Analysis of 500 Novel Complete Protein Coding Human cDNAs.";
RL   Genome Res. 11:422-435(2001).
DR   EMBL; AL122042; CAB59179.2; -.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR01217; PRICHEXTENSN.
KW   Hypothetical protein.
SQ   SEQUENCE   148 AA;  15375 MW;  796D11BA16A0591E CRC64;

  Query Match             7.7%;  Score 153.5;  DB 4;  Length 148;
  Best Local Similarity   30.4%;  Pred. No. 0.011;
  Matches   42;  Conservative   18;  Mismatches   45;  Indels   33;  Gaps    6;

Qy        126 GDEPPVFSRPISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNP---SGYPGC---- 178
              |   |    ||:  :||     |      || | |   :||| ||:|    |||||
Db         22 GGSNPAHPPPINPPFPPGPCPPP------PGAPHGNPAFPSGGPPHPVPQPGYPGCQPLG 75

Qy        179 PYPPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDR 238
              . ||||  | ||      |  |||  : |   |            : : ||   :|:::| :
Db         76 PYPPPYPPPA------PGIPPVNPLAPGMVGP----------AVIVDK---KMQKKMKK 115

Qy        239 AQAELNALKRTEEDLKKG 256
              |  :::  ::  :  | |
Db        116 AHKKMHKHQKHHKYHKHG 133


RESULT 15
Q9NZ81
ID   Q9NZ81      PRELIMINARY;      PRT;   148 AA.
AC   Q9NZ81;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Uncharacterized bone marrow protein BM041 (DKFZp564J157 protein)
DE   (Similar to DKFZp564J157 protein).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BONE MARROW;
RA   Zhao M., Gu J., Li N., Peng Y., Han Z., Chen Z.;
RT   "A novel gene expressed in human bone marrow.";
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=B-CELL;
RA   Strausberg R.;
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=B-CELL;
RA   Strausberg R.;
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF217517; AAF67628.1; -.
DR   EMBL; BC016064; AAH16064.1; -.
DR   EMBL; BC014257; AAH14257.1; -.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR01217; PRICHEXTENSN.
SQ   SEQUENCE   148 AA;  15385 MW;  56BFEE6A15935A2E CRC64;

  Query Match             7.4%;  Score 148.5;  DB 4;  Length 148;
  Best Local Similarity   29.7%;  Pred. No. 0.023;
  Matches   41;  Conservative   18;  Mismatches   46;  Indels   33;  Gaps    6;

Qy        126 GDEPPVFSRPISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNP---SGYPGC---- 178
              |   |    ||:  :||     |      || | |   :| | ||:|    |||||
Db         22 GGSNPAHPPPINPPFPPGPCPPP------PGAPHGNPAFPPGGPPHPVPQPGYPGCQPLG 75

Qy        179 PYPPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDR 238
              ||||  | ||      |  |||  : |   |            : : ||   :|:::| :
Db         76 PYPPPYPPPA------PGIPPVNPLAPGMVGP----------AVIVDK---KMQKKMKK 115
```

```
Qy        239 AQAELNALKRTEEDLKKG 256
              |  :::  ::  :  | |
Db        116 AHKKMHKHQKHHKYHKHG 133
```

Search completed: June 3, 2003, 14:11:51
Job time : 53.9304 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: June 3, 2003, 14:05:55 ; Search time 51.5677 Seconds
(without alignments)
984.502 Million cell updates/sec

Title: US-09-804-690-2
Perfect score: 1996
Sequence: 1 MMSKYKYRDLTVRQTVNVIA.....FQLRALMQKARKTAGLSDLY 381

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1996	100.0	381	18	AAW19110	Mouse tumour susce
2	1996	100.0	381	20	AAW93425	Mouse tsg101 prote
3	1900.5	95.2	380	18	AAW19111	Human tumour susce
4	1900.5	95.2	390	20	AAW93424	Human NHTS protein
5	1337.5	67.0	307	23	ABP41729	Human ovarian anti
6	1022.5	51.2	237	22	AAE09328	Human intracellula
7	786	39.4	331	22	ABB64607	Drosophila melanog
8	440	22.0	398	21	AAG06370	Arabidopsis thalia
9	440	22.0	412	21	AAG06369	Arabidopsis thalia
10	406.5	20.4	340	21	AAG06371	Arabidopsis thalia
11	402	20.1	90	21	AAG01689	Human secreted pro
12	364.5	18.3	322	21	AAG13898	Arabidopsis thalia
13	361	18.1	341	22	AAB93473	Human protein sequ
14	327	16.4	146	21	AAB53717	Human colon cancer
15	318	15.9	452	22	AAU23116	Novel human enzyme
16	266	13.3	288	21	AAG13899	Arabidopsis thalia
17	162.5	8.1	168	21	AAG13900	Arabidopsis thalia
18	152	7.6	218	22	ABG18833	Novel human diagno
19	152	7.6	485	21	AAB57086	Human prostate can
20	146	7.3	148	21	AAY86515	Human gene 71-enco
21	146	7.3	148	23	ABB97431	Novel human protei
22	146	7.3	176	22	AAG75132	Human colon cancer
23	143	7.2	777	23	ABP41839	Human ovarian anti
24	142.5	7.1	905	18	AAW31186	Human p160 polypep
25	142.5	7.1	1135	18	AAW31185	Human p160 polypep
26	142	7.1	466	11	AAR07084	Recombinant human
27	138	6.9	258	21	AAB44500	Plant viral moveme
28	137.5	6.9	244	21	AAG40015	Arabidopsis thalia
29	137.5	6.9	245	21	AAG05889	Arabidopsis thalia
30	137.5	6.9	245	21	AAG38059	Arabidopsis thalia
31	137.5	6.9	247	21	AAG05888	Arabidopsis thalia
32	137.5	6.9	247	21	AAG38058	Arabidopsis thalia
33	137.5	6.9	263	21	AAG38057	Arabidopsis thalia
34	137.5	6.9	274	21	AAG05887	Arabidopsis thalia
35	137	6.9	397	22	AAM39852	Human polypeptide
36	137	6.9	397	22	AAG81381	Human AFP protein
37	137	6.9	442	22	AAM41638	Human polypeptide
38	135.5	6.8	827	23	ABP35602	Fungal ZBC protein
39	135	6.8	172	22	ABG18832	Novel human diagno
40	133	6.7	505	19	AAW53808	N-Wiskott-Aldrich
41	133	6.7	505	22	AAM52320	Bovine N-WASP prot
42	133	6.7	505	22	AAG67360	Amino acid sequenc
43	128	6.4	505	20	AAY07117	Lung cancer associ
44	126.5	6.3	505	19	AAW46889	Human Neural-Wisko
45	126.5	6.3	505	22	AAM52316	Human N-WASP prote

ALIGNMENTS

RESULT 1
AAW19110
ID AAW19110 standard; Protein; 381 AA.
XX
AC AAW19110;
XX
DT 10-DEC-1997 (first entry)
XX
DE Mouse tumour susceptibility protein TSG101.
XX
KW Tumour susceptibility protein; TSG101; tumorigenesis;
KW breast cancer; diagnosis; therapy; mouse.
XX
OS Mus musculus.
XX
FH Key Location/Qualifiers
FT Region 37..46
FT /note= "residues 37-46 resemble a helix-turn-helix
FT signature domain"
FT Region 73..83
FT /note= "residues 73-83 resemble a fungal Zn-cys
FT bi-nuclear cluster signature "
FT Modified-site 11
FT /note= "potential protein kinase C phosphorylation
FT site"
FT Modified-site 38
FT /note= "potential protein kinase C phosphorylation
FT site"
FT Modified-site 85
FT /note= "potential protein kinase C phosphorylation
FT site"
FT Modified-site 88
FT /note= "potential protein kinase C phosphorylation

```
FT                              site"
FT   Modified-site   215
FT                   /note= "potential protein kinase C phosphorylation
FT                          site"
FT   Modified-site   225
FT                   /note= "potential protein kinase C phosphorylation
FT                          site"
FT   Modified-site   357
FT                   /note= "potential protein kinase C phosphorylation
FT                          site"
FT   Modified-site   38
FT                   /note= "potential casein kinase II phosphorylation
FT                          site"
FT   Modified-site   210
FT                   /note= "potential casein kinase II phosphorylation
FT                          site"
FT   Modified-site   249
FT                   /note= "potential casein kinase II phosphorylation
FT                          site"
FT   Modified-site   265
FT                   /note= "potential casein kinase II phosphorylation
FT                          site"
FT   Modified-site   290
FT                   /note= "potential casein kinase II phosphorylation
FT                          site"
FT   Modified-site   55
FT                   /note= "potential N-myristoylation site"
FT   Modified-site   156
FT                   /note= "potential N-myristoylation site"
FT   Modified-site   44
FT                   /note= "potential N-glycosylation site"
FT   Modified-site   150
FT                   /note= "potential N-glycosylation site"
FT   Modified-site   297
FT                   /note= "potential N-glycosylation site"
XX
PN   WO9718333-A1.
XX
PD   22-MAY-1997.
XX
PF   15-NOV-1996;   96WO-US18828.
XX
PR   13-JUN-1996;   96US-0670274.
PR   16-NOV-1995;   95US-0006856.
PR   16-JAN-1996;   96US-0585758.
XX
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
XX
PI   Cohen SN,  Li L;
XX
DR   WPI; 1997-289308/26.
DR   N-PSDB; AAT69574.
XX
PT   Nucleic acid encoding tumour susceptibility protein TSG101 - useful
PT   to develop products for cancer detection, prophylaxis, therapy and
PT   predisposition determination
XX
PS   Example 1; Page 67-70; 89pp; English.
XX
CC   This polypeptide sequence comprises a mouse tumour susceptibility
CC   protein, designated TSG101, that can function as a tumour
CC   suppressor in oncogenesis.  Its amino acid sequence was deduced
CC   from a cDNA clone (AAT69574) isolated using a novel random homozygous
CC   knockout approach.  It shows 94% identity to human TSG101 (see
CC   AAW19111).  Deletions in human TSG101 are associated with the
CC   occurrence of cancers, e.g. breast cancer.  Methods are provided
CC   for identifying homologous or related proteins, for producing
CC   compositions that modulate the expression or function of the TSG101
CC   protein, and for studying physiological pathways.  Modulation of
CC   gene activity in vivo is used for prophylactic or therapeutic
CC   purposes, such as cancer therapy, identification of cell type based
CC   on expression, etc.
XX
```

```
SQ   Sequence   381 AA;

  Query Match             100.0%;  Score 1996;  DB 18;  Length 381;
  Best Local Similarity   100.0%;  Pred. No. 5.5e-146;
  Matches  381;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNI 60

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQI 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQI 120

Qy        121 MIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCP 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 MIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCP 180

Qy        181 YPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 YPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGA 240

Qy        241 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 300

Qy        301 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 360

Qy        361 QFQLRALMQKARKTAGLSDLY 381
              |||||||||||||||||||||
Db        361 QFQLRALMQKARKTAGLSDLY 381
```

```
RESULT 2
AAW93425
ID   AAW93425 standard; Protein; 381 AA.
XX
AC   AAW93425;
XX
DT   11-JUN-1999  (first entry)
XX
DE   Mouse tsg101 protein.
XX
KW   Human tumour suppressor; NHTS; treatment; diagnosis; lymphoma; cancer;
KW   brain; breast; colon; heart; kidney; ovary; paraganglia; pancreas;
KW   prostate; skin; stomach; thyroid; autoimmune disease; asthma; diabetes;
KW   biliary cirrhosis; Crohn's disease; rheumatoid arthritis; mouse; tsg101.
XX
OS   Mus sp.
XX
PN   US5892016-A.
XX
PD   06-APR-1999.
XX
PF   23-JAN-1997;   97US-0786999.
XX
PR   23-JAN-1997;   97US-0786999.
XX
PA   (INCY-) INCYTE PHARM.
XX
PI   Brie SL,  Goli SK;
XX
DR   WPI; 1999-253932/21.
XX
PT   Novel human tumor suppressor - useful for the diagnosis or treatment
PT   of lymphoma, cancer, and autoimmune disease
XX
PS   Disclosure; Fig 2; 25pp; English.
XX
CC   This invention describes the isolation of a novel human tumour
CC   suppressor (NHTS). The products of the invention may be used for
```

```
CC   the diagnosis or treatment of conditions and diseases which are
CC   associated with expression of NHTS e.g. lymphoma and cancers of the
CC   brain, breast, colon, heart, kidney, ovary, paraganglia, pancreas,
CC   prostate, skin, stomach and thyroid and autoimmune disease e.g. asthma,
CC   biliary cirrhosis, Crohn's disease, diabetes, and rheumatoid arthritis.
XX
SQ   Sequence   381 AA;

  Query Match           100.0%;  Score 1996;  DB 20;  Length 381;
  Best Local Similarity  100.0%;  Pred. No. 5.5e-146;
  Matches  381;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNI 60

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQI 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQI 120

Qy        121 MIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCP 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 MIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCP 180

Qy        181 YPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 YPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGA 240

Qy        241 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 300

Qy        301 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 360

Qy        361 QFQLRALMQKARKTAGLSDLY 381
              |||||||||||||||||||||
Db        361 QFQLRALMQKARKTAGLSDLY 381


RESULT 3
AAW19111
ID   AAW19111 standard; Protein; 380 AA.
XX
AC   AAW19111;
XX
DT   10-DEC-1997  (first entry)
XX
DE   Human tumour susceptibility protein TSG101.
XX
KW   Tumour susceptibility protein; TSG101; tumorigenesis;
KW   breast cancer; diagnosis; therapy; human.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          130..205
FT                   /note= "proline-rich domain"
FT   Domain          231..302
FT                   /note= "coiled-coil domain"
FT   Modified-site   11
FT                   /note= "potential protein kinase C phosphorylation
FT                    site"
FT   Modified-site   38
FT                   /note= "potential protein kinase C phosphorylation
FT                    site"
FT   Modified-site   86
FT                   /note= "potential protein kinase C phosphorylation
FT                    site"
FT   Modified-site   89
FT                   /note= "potential protein kinase C phosphorylation
FT                    site"
FT   Modified-site   215
FT                   /note= "potential protein kinase C phosphorylation
FT                    site"
FT   Modified-site   225
FT                   /note= "potential protein kinase C phosphorylation
FT                    site"
FT   Modified-site   357
FT                   /note= "potential protein kinase C phosphorylation
FT                    site"
FT   Modified-site   38
FT                   /note= "potential casein kinase II phosphorylation
FT                    site"
FT   Modified-site   210
FT                   /note= "potential casein kinase II phosphorylation
FT                    site"
FT   Modified-site   249
FT                   /note= "potential casein kinase II phosphorylation
FT                    site"
FT   Modified-site   265
FT                   /note= "potential casein kinase II phosphorylation
FT                    site"
FT   Modified-site   290
FT                   /note= "potential casein kinase II phosphorylation
FT                    site"
FT   Modified-site   44
FT                   /note= "potential N-glycosylation site"
FT   Modified-site   150
FT                   /note= "potential N-glycosylation site"
FT   Modified-site   297
FT                   /note= "potential N-glycosylation site"
XX
PN   WO9718333-A1.
XX
PD   22-MAY-1997.
XX
PF   15-NOV-1996;   96WO-US18828.
XX
PR   13-JUN-1996;   96US-0670274.
PR   16-NOV-1995;   95US-0006856.
PR   16-JAN-1996;   96US-0585758.
XX
PA   (STRD ) UNIV LELAND STANFORD JUNIOR.
XX
PI   Cohen SN,  Li L;
XX
DR   WPI; 1997-289308/26.
DR   N-PSDB; AAT69575.
XX
PT   Nucleic acid encoding tumour susceptibility protein TSG101 - useful
PT   to develop products for cancer detection, prophylaxis, therapy and
PT   predisposition determination
XX
PS   Example 1; Page 73-74; 89pp; English.
XX
CC   This polypeptide sequence comprises a human tumour susceptibility
CC   protein, designated TSG101, that can function as a tumour
CC   suppressor in oncogenesis.  Its amino acid sequence was deduced
CC   from a cDNA clone (AAT69575) isolated from a human placenta cDNA
CC   library.  It shows 94% identity to human TSG101 (see AAW19110).
CC   Deletions in human TSG101 are associated with the occurrence of
CC   cancers, e.g. breast cancer.  Methods are provided for identifying
CC   homologous or related proteins, for producing compositions that
CC   modulate the expression or function of the TSG101 protein, and for
CC   studying physiological pathways.  Modulation of gene activity in
CC   vivo is used for prophylactic or therapeutic purposes, such as
CC   cancer therapy, identification of cell type based on expression,
CC   etc.
XX
SQ   Sequence   380 AA;

  Query Match           95.2%;  Score 1900.5;  DB 18;  Length 380;
  Best Local Similarity  94.5%;  Pred. No. 1.3e-138;
  Matches  360;  Conservative   11;  Mismatches    9;  Indels    1;  Gaps    1;
```

```
Qy          1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNI 60
              |:||||||||||:||||| :|||||||||||||||||||:||||||| |||| |||
Db          1 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 60

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQI 120
              ||||||||||||||||||||||||||||||||||||||||:||||:|:|| |||:
Db         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120

Qy        121 MIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCP 180
              |||:||:|||||||| :||||||| |||||||||||| ||| ||||||||||||||
Db        121 MIVVFGDEPPVFSRP-ISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCP 179

Qy        181 YPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGA 240
              ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        180 YPPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRA 239

Qy        241 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        240 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 299

Qy        301 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        300 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 359

Qy        361 QFQLRALMQKARKTAGLSDLY 381
              |||||||||||||||||||||
Db        360 QFQLRALMQKARKTAGLSDLY 380
```

RESULT 4
AAW93424
ID AAW93424 standard; Protein; 390 AA.
XX
AC AAW93424;
XX
DT 11-JUN-1999 (first entry)
XX
DE Human NHTS protein.
XX
KW Human tumour suppressor; NHTS; treatment; diagnosis; lymphoma; cancer;
KW brain; breast; colon; heart; kidney; ovary; paraganglia; pancreas;
KW prostate; skin; stomach; thyroid; autoimmune disease; asthma; diabetes;
KW biliary cirrhosis; Crohn's disease; rheumatoid arthritis.
XX
OS Homo sapiens.
XX
PN US5892016-A.
XX
PD 06-APR-1999.
XX
PF 23-JAN-1997; 97US-0786999.
XX
PR 23-JAN-1997; 97US-0786999.
XX
PA (INCY-) INCYTE PHARM.
XX
PI Brie SL, Goli SK;
XX
DR WPI; 1999-253932/21.
DR N-PSDB; AAX23168.
XX
PT Novel human tumor suppressor - useful for the diagnosis or treatment
PT of lymphoma, cancer, and autoimmune disease
XX
PS Claim 1; Fig 1A-B; 25pp; English.
XX
CC This invention describes the isolation of a novel human tumour
CC suppressor (NHTS). The products of the invention may be used for
CC the diagnosis or treatment of conditions and diseases which are
CC associated with expression of NHTS e.g. lymphoma and cancers of the
CC brain, breast, colon, heart, kidney, ovary, paraganglia, pancreas,
CC prostate, skin, stomach and thyroid and autoimmune disease e.g. asthma,
CC biliary cirrhosis, Crohn's disease, diabetes, and rheumatoid arthritis.
XX
SQ Sequence 390 AA;

Query Match 95.2%; Score 1900.5; DB 20; Length 390;
Best Local Similarity 94.5%; Pred. No. 1.4e-138;
Matches 360; Conservative 11; Mismatches 9; Indels 1; Gaps 1;

```
Qy          1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNI 60
              |:||||||||||:||||| :|||||||||||||||||||:||||||| |||| |||
Db         11 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 70

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQI 120
              ||||||||||||||||||||||||||||||||||||||||:||||:|:|| |||:
Db         71 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 130

Qy        121 MIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCP 180
              |||:||:|||||||| :||||||| |||||||||||| ||| ||||||||||||||
Db        131 MIVVFGDEPPVFSRP-ISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCP 189

Qy        181 YPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGA 240
              ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        190 YPPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRA 249

Qy        241 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        250 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 309

Qy        301 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        310 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 369

Qy        361 QFQLRALMQKARKTAGLSDLY 381
              |||||||||||||||||||||
Db        370 QFQLRALMQKARKTAGLSDLY 390
```

RESULT 5
ABP41729
ID ABP41729 standard; Protein; 307 AA.
XX
AC ABP41729;
XX
DT 22-AUG-2002 (first entry)
XX
DE Human ovarian antigen HSPSE88, SEQ ID NO:2861.
XX
KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
KW ovarian cancer; breast cancer; tumour; reproductive system disorder;
KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
KW PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;
KW inflammatory condition; immune disorder; blood disorder;
KW cardiovascular disorder; respiratory disorder; neurological disorder;
KW gastrointestinal disorder; urinary system disorder; drug screening;
KW gene therapy; chromosome mapping; forensic analysis;
KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
KW antiinflammatory; gynaecological; reproductive; chromosome 11p15.
XX
OS Homo sapiens.
XX
PN WO200200677-A1.
XX
PD 03-JAN-2002.
XX
PF 07-JUN-2001; 2001WO-US18569.
XX
PR 07-JUN-2000; 2000US-209467P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Birse CE, Rosen CA;
XX
DR WPI; 2002-147878/19.

```
DR   N-PSDB; ABQ54806.
XX
PT   Isolated nucleic acid molecules encoding novel ovarian polypeptides,
PT   useful in the prevention, treatment and diagnosis of cancer (e.g.
PT   ovarian cancer), immune disorders, cardiovascular disorders and
PT   neurological diseases -
XX
PS   Claim 11; SEQ ID No 2861; 2922pp; English.
XX
CC   The invention relates to 2175 novel human ovarian antigens (ABP41054-
CC   ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
CC   encompasses polypeptides 90% identical and polynucleotides 95% identical
CC   to the sequences of the invention. The invention additionally relates to
CC   recombinant vectors and host cells comprising human ovarian antigen
CC   polynucleotides, antibodies against human ovarian antigens, and the use
CC   of ovarian antigen polynucleotides and polypeptides in diagnosing,
CC   treating, prognosing or preventing various ovary and/or breast-related
CC   disorders. Such conditions include ovarian cancer and breast cancer, and
CC   metastatic tumours of ovarian or breast origin, reproductive system
CC   disorders (e.g., infertility, disorders of pregnancy, anovulation,
CC   polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
CC   disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
CC   shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
CC   vaginitis), immune disorders (e.g., congenital and acquired
CC   immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
CC   blood-related disorders (e.g., anaemia), cardiovascular disorders,
CC   respiratory disorders, neurological disorders, gastrointestinal disorders
CC   and urinary system disorders. Ovarian antigen polypeptides and
CC   polynucleotides may also be used in screening for compounds which
CC   modulate ovarian antigen expression or activity. The polynucleotides may
CC   further be used for gene therapy, chromosome mapping, in the
CC   identification of individuals and in forensic analysis, and the
CC   polypeptides may be used as food additives or to prepare antibodies
CC   useful in disease diagnosis, drug targeting and phenotyping. The present
CC   sequence represents a human ovarian antigen of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   307 AA;

  Query Match           67.0%;  Score 1337.5;  DB 23;  Length 307;
  Best Local Similarity   95.2%;  Pred. No. 3.2e-95;
  Matches  259;  Conservative    6;  Mismatches    6;  Indels    1;  Gaps    1;

Qy        110 PRSELLELIQIMIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGY 169
              |:|:|| |||:|||:||:|||||||| :||||||| ||||||||||||||| ||| |||||
Db         37 PQSDLLGLIQVMIVVFGDEPPVFSRP-ISASYPPYQATGPPNTSYMPGMPGGISPYPSGY 95

Qy        170 PPNPSGYPGCPYPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKL 229
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db         96 PPNPSGYPGCPYPPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKL 155

Qy        230 RWRMKEEMDGAQAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEEL 289
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        156 RWRMKEEMDRAQAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEEL 215

Qy        290 SSALEKMENQSENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDV 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        216 SSALEKMENQSENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDV 275

Qy        350 FLKHVRLLSRKQFQLRALMQKARKTAGLSDLY 381
              |||||||||||||||||||||||||||||||
Db        276 FLKHVRLLSRKQFQLRALMQKARKTAGLSDLY 307


RESULT 6
AAE09328
ID   AAE09328 standard; Protein; 237 AA.
XX
AC   AAE09328;
XX
DT   19-NOV-2001  (first entry)
XX
DE   Human intracellular regulatory molecule, tsg101.
XX
KW   Human; intracellular regulator; cell division; proliferation; therapy;
KW   cancer; infection; wound; developmental abnormality; metabolic problem;
KW   cytostatic; antibacterial; vulnerary; transcription factor; tsg101.
XX
OS   Homo sapiens.
XX
PN   US6274312-B1.
XX
PD   14-AUG-2001.
XX
PF   10-DEC-1997;   97US-0999774.
XX
PR   11-DEC-1996;   96US-0032818.
XX
PA   (SCHE ) SCHERING CORP.
XX
PI   Gish KC,  Seghezzi W,  Shanahan F,  Lees EM,  McClanahan TK;
XX
DR   WPI; 2001-535086/59.
DR   N-PSDB; AAD16104.
XX
PT   New genes encoding intracellular regulatory molecules, useful for
PT   regulating cell division and proliferation (e.g. tumor cells),
PT   particularly for treating cancer, infections, wounds, or developmental
PT   or metabolic abnormalities -
XX
PS   Example 2; Column 29-32; 62pp; English.
XX
CC   The present sequence is a human intracellular regulatory
CC   molecule, tsg101. The tsg101 is a transcription factor. The
CC   polynucleotides encoding intracellular regulatory molecules are
CC   useful for regulating cell division and proliferation of various
CC   cell types, including tumour cells. Specifically, they are also
CC   useful for treating cancer, infections, wounds, developmental
CC   abnormalities or metabolic problems.
XX
SQ   Sequence   237 AA;

  Query Match           51.2%;  Score 1022.5;  DB 22;  Length 237;
  Best Local Similarity   88.3%;  Pred. No. 5e-71;
  Matches  182;  Conservative   12;  Mismatches   11;  Indels    1;  Gaps    1;

Qy          1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNI 60
              |:|||||||||||:|||| :|||||||||||||||||||:|||||||| |||| |||
Db         11 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 70

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQI 120
              ||||||||||||||||||||||||||||||||||||||||||||||:||||:|:|| |||:
Db         71 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 130

Qy        121 MIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCP 180
              |||:||:||||||||| :||||||| |||||||||||||| ||| ||||||||||||||||
Db        131 MIVVFGDEPPVFSRP-ISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCP 189

Qy        181 YPPAGPYPATTSSQYPSQPPVTTVGP 206
              ||| |||||||||||||||  : |
Db        190 YPPGGPYPATTSSQYPSQPPVPLLVP 215


RESULT 7
ABB64607
ID   ABB64607 standard; Protein; 331 AA.
XX
AC   ABB64607;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 20613.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
```

```
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL08710.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 20613; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   331 AA;

  Query Match             39.4%;  Score 786;  DB 22;  Length 331;
  Best Local Similarity   49.0%;  Pred. No. 1.5e-52;
  Matches  166;  Conservative   55;  Mismatches   82;  Indels   36;  Gaps    9;

Qy         67 LDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQIMIVIFG 126
              :||:| | |:|||||| :| ||   :|| |||:||||||||:   |:|| |||:||| ||
Db          1 MDTHPQNAPMCFVKPTPTMQIKVSMYVDHNGKVYLPYLHDWQPHSSDLLSLIQVMIVTFG 60

Qy        127 EEPPVFSRP--TVSASYPPYTATGPPNTSYM--PGMPSGISA---YP-------SGYPPN 172
              : |||:|:|   ::| |        |  |||  || | | ::   ||'      | :||
Db         61 DHPPVYSKPKEQIAAPY--------PTNSYMPQPGAPGGSNSFLPYPTAGGAGGSNFPPY 112

Qy        173 PSGYPGCPYP--PAGP-----YPA------TTSSQYPSQPPVTTVGPSRDGTISEDTIRA 219
              |:|    |||  ||||     ||       |:  ||          ||  |||:|: |:|
Db        113 PTGSNVGPYPPTPAGPAGGSGYPAYPNFIQPTAGGYPPAAGYNPSNPSSTGTITEEHIKA 172

Qy        220 SLISAVSDKLRWRMKEEMDGAQAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNI 279
              |:|||: |||| |::|:::  |||:  | ||:::| :|  |:: :: ||::|  :: |||
Db        173 SIISAIDDKLRRRVQEKVNQYQAEIETLNRTKQELLEGSAKIDAIIERLEREHIDMQKNI 232

Qy        280 ELLKKKDEELSSALEKMENQSENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEA 339
               :|| |::||  ||| :|:    |  || :  |||||:|:|| ||:| | || |:||||
Db        233 SILKDKEQELEKALEDLESAEAINP-DEAVTTTAPLYRQLLNAYADEAATEDAIYYLGEG 291

Qy        340 LRRGVIDLLDVFLKHVRLLSRKQFQLRALMQKARKTAGLS 378
              || |||||: |||||| |||||| ||| ||| |: |||:
Db        292 LRGGVIDLETFLKHVRQLSRKQFILRATMQKCRQKAGLA 330


RESULT 8
AAG06370
ID   AAG06370 standard; Protein; 398 AA.
XX
```

```
AC   AAG06370;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 3119.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
```

PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 22.0%; Score 440; DB 21; Length 398;
Best Local Similarity 31.6%; Pred. No. 1e-25;
Matches 117; Conservative 72; Mismatches 147; Indels 34; Gaps 12;

```
Qy         12 VRQ-TVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNIPICLWLLDTY 70
              :||  :|:|: |  |:|   |:: ||| |  |:   ||||: : |  ||||: :|||::|
Db         37 IRQHLLNLISSYPSLEPKTASFMHNDGRSVNLLQADGTIPMPFHGVTYNIPVIIWLLESY 96

Qy         71 PYNPPICFVKPTSSMTIK-TGKHVDANGKIYLPYLHDWKHPRSELLELIQIMIVIFGEEP 129
              |.:||  :| ||: | ||    ||  :| : |||| :| :| | |::|:  :   |  :|
Db         97 PRHPPCVYVNPTADMIIKRPHAHVTPSGLVSLPYLQNWVYPSSNLVDLVSDLSAAFARDP 156

Qy        130 PVFSRPTVSASYPPYTATGPPNT--SYMPGMPSGISAYPSGYPPNPSGYPGCPYPPAGPY 187
              |::||       ||     ||    | :   ||   : |  :||:| |         |
Db        157 PLYSR---RRPQPP--PPSPPTVYDSSLSRPPSADQSLPRPFPPSPYG--------GGVS 203

Qy        188 PATTSSQYPSQPPVTTVGPSRDGTISE--DTIRASLISAVSDKLRWRMKEEMDGAQAELN 245
                :  |       :  |::  : : : |:|    : |    ||:   ||
Db        204 RVQVQHVHHQQQSDDAAEVFKRNAINKMVEMVHSDLVS--MRRAREAEAEELLSLQA--- 258
```

```
Qy        246 ALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQSEN--- 302
              ||| |::|  |   |:|||    :|   ::: ::::    : | |.: : : :::|
Db        259 GLKRREDELNIG---LKEMV----EEKETLEQQLQIISMNTDILDSWVRENQGKTKNLVD 311

Qy        303 NDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRKQF 362
              |:|        | ||:|   | : |||| |: | :: : ||:  | :|::||||||:|| 
Db        312 LDVDNAFECGDTLSKQMLECTALDLAIEDAIYSLDKSFQDGVVPFDQYLRNVRLLSREQF 371

Qy        363 QLRALMQKAR 372
              ||   | |
Db        372 FHRATGSKVR 381


RESULT 9
AAG06369
ID   AAG06369 standard; Protein; 412 AA.
XX
AC   AAG06369;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 3118.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145218.
PR   23-JUL-1999;   99US-0145224.
PR   26-JUL-1999;   99US-0145276.
PR   27-JUL-1999;   99US-0145913.
PR   27-JUL-1999;   99US-0145918.
PR   27-JUL-1999;   99US-0145919.
PR   28-JUL-1999;   99US-0145951.
PR   02-AUG-1999;   99US-0146386.
PR   02-AUG-1999;   99US-0146388.
PR   02-AUG-1999;   99US-0146389.
PR   03-AUG-1999;   99US-0147038.
PR   04-AUG-1999;   99US-0147204.
PR   04-AUG-1999;   99US-0147302.
PR   05-AUG-1999;   99US-0147192.
PR   05-AUG-1999;   99US-0147260.
PR   06-AUG-1999;   99US-0147303.
```

```
PR   06-AUG-1999;   99US-0147416.
PR   09-AUG-1999;   99US-0147493.
PR   09-AUG-1999;   99US-0147935.
PR   10-AUG-1999;   99US-0148171.
PR   11-AUG-1999;   99US-0148319.
PR   12-AUG-1999;   99US-0148341.
PR   13-AUG-1999;   99US-0148565.
PR   13-AUG-1999;   99US-0148684.
PR   16-AUG-1999;   99US-0149368.
PR   17-AUG-1999;   99US-0149175.
PR   18-AUG-1999;   99US-0149426.
PR   20-AUG-1999;   99US-0149722.
PR   20-AUG-1999;   99US-0149723.
PR   20-AUG-1999;   99US-0149929.
PR   23-AUG-1999;   99US-0149902.
PR   23-AUG-1999;   99US-0149930.
PR   25-AUG-1999;   99US-0150566.
PR   26-AUG-1999;   99US-0150884.
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

 Query Match             22.0%;  Score 440;  DB 21;  Length 412;
 Best Local Similarity   31.6%;  Pred. No. 1.1e-25;
 Matches  117;  Conservative   72;  Mismatches  147;  Indels   34;  Gaps   12;

Qy          12 VRQ-TVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNIPICLWLLDTY 70
               :||  :|:|: |  |:|   |:: ||| |  |:   ||||: : |  ||||: :|||::|
Db          51 IRQHLLNLISSYPSLEPKTASFMHNDGRSVNLLQADGTIPMPFHGVTYNIPVIIWLLESY 110

Qy          71 PYNPPICFVKPTSSMTIK-TGKHVDANGKIYLPYLHDWKHPRSELLELIQIMIVIFGEEP 129
               | :||  :| ||: | ||    ||  :| : |||| :| :| | |::|:  :   | :|
Db         111 PRHPPCVYVNPTADMIIKRPHAHVTPSGLVSLPYLQNWVYPSSNLVDLVSDLSAAFARDP 170

Qy         130 PVFSRPTVSASYPPYTATGPPNT--SYMPGMPSGISAYPSGYPPNPSGYPGCPYPPAGPY 187
               |::||       ||     ||    | :   ||   : |  :||:| |         |
Db         171 PLYSR---RRPQPP--PPSPPTVYDSSLSRPPSADQSLPRPFPPSPYG--------GGVS 217

Qy         188 PATTSSQYPSQPPVTTVGPSRDGTISE--DTIRASLISAVSDKLRWRMKEEMDGAQAELN 245
                    :  |          :   |::  : : : |:|    : |    ||:   ||
Db         218 RVQVQHVHHQQQSDDAAEVFKRNAINKMVEMVHSDLVS--MRRAREAEAEELLSLQA--- 272

Qy         246 ALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQSEN--- 302
               ||| |::|  |   |:|||    :|   ::: ::::    : | | : : : :::|
Db         273 GLKRREDELNIG---LKEMV----EEKETLEQQLQIISMNTDILDSWVRENQGKTKNLVD 325

Qy         303 NDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRKQF 362
               |:|         | ||:|   | : |||| |: | :: : ||:  | :|::|||||||:||
Db         326 LDVDNAFECGDTLSKQMLECTALDLAIEDAIYSLDKSFQDGVVPFDQYLRNVRLLSREQF 385

Qy         363 QLRALMQKAR 372
                 ||  | |
Db         386 FHRATGSKVR 395


RESULT 10
AAG06371
ID   AAG06371 standard; Protein; 340 AA.
XX
AC   AAG06371;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 3120.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
```

PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.

```
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match            20.4%;  Score 406.5;  DB 21;  Length 340;
  Best Local Similarity  31.5%;  Pred. No. 3.3e-23;
  Matches  109;  Conservative   65;  Mismatches  139;  Indels   33;  Gaps   11;

Qy         35 NDGSSRELVNLTGTIPVRYRGNIYNIPICLWLLDTYPYNPPICFVKPTSSMTIK-TGKHV 93
              ||| |  |:   ||||: : |  ||||: :|||::|| :||  :| ||: | ||    ||
Db          3 NDGRSVNLLQADGTIPMPFHGVTYNIPVIIWLLESYPRHPPCVYVNPTADMIIKRPHAHV 62

Qy         94 DANGKIYLPYLHDWKHPRSELLELIQIMIVIFGEEPPVFSRPTVSASYPPYTATGPPNT- 152
               :| : |||| :| :| | |::|:  :   |  :||::||       ||     ||
Db         63 TPSGLVSLPYLQNWVYPSSNLVDLVSDLSAAFARDPPLYSR---RRPQPP--PPSPPTVY 117

Qy        153 -SYMPGMPSGISAYPSGYPPNPSGYPGCPYPPAGPYPATTSSQYPSQPPVTTVGPSRDGT 211
               | :   ||   : |  :||:| |          |       :  |       :
Db        118 DSSLSRPPSADQSLPRPFPPSPYG--------GGVSRVQVQHVHHQQQSDDAAEVFKRNA 169

Qy        212 ISE--DTIRASLISAVSDKLRWRMKEEMDGAQAELNALKRTEEDLKKGHQKLEEMVTRLD 269
              |::  : : : |:|    : |    ||:   ||    ||| |::|  |   |:|||
Db        170 INKMVEMVHSDLVS--MRRAREAEAEELLSLQA---GLKRREDELNIG---LKEMV---- 217

Qy        270 QEVAEVDKNIELLKKKDEELSSALEKMENQSEN---NDIDEVIIPTAPLYKQILNLYAEE 326
              :|   ::: ::::    : | | : : : :::|    |:|        | ||:|   | :
Db        218 EEKETLEQQLQIISMNTDILDSWVRENQGKTKNLVDLDVDNAFECGDTLSKQMLECTALD 277

Qy        327 NAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRKQFQLRALMQKAR 372
              |||| |: | :: : ||:  | :|::||||||:||  ||   | |
Db        278 LAIEDAIYSLDKSFQDGVVPFDQYLRNVRLLSREQFFHRATGSKVR 323


RESULT 11
AAG01689
ID   AAG01689 standard; Protein; 90 AA.
XX
AC   AAG01689;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein, SEQ ID NO: 5770.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   N-PSDB; AAC01695.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 13; SEQ ID 5770; 71pp + CD-ROM; English.
XX
CC   The present sequence is a polypeptide encoded by one of a large number
CC   of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC   were prepared from total human RNAs or polyA+ RNAs derived from 30
CC   different tissues. EST sequences usually correspond mainly to the 3'
CC   untranslated region (UTR) of the mRNA because they are often obtained
CC   from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC   isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC   those cases where longer cDNA sequences have been obtained, the full 5'
CC   UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC   ends and can therefore be used to obtain full length cDNAs and genomic
CC   DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC   chromosome mapping procedures. They are used to obtain upstream
CC   regulatory sequences and to design expression and secretion vectors.
XX
SQ   Sequence   90 AA;

  Query Match            20.1%;  Score 402;  DB 21;  Length 90;
  Best Local Similarity  91.2%;  Pred. No. 1.2e-23;
  Matches   73;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNI 60
              |:||||||||||:||||| :|||||||||||||||||||:|||||||| |||| |||
Db         11 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 70

Qy         61 PICLWLLDTYPYNPPICFVK 80
              ||||||||||||||||||||
Db         71 PICLWLLDTYPYNPPICFVK 90


RESULT 12
AAG13898
ID   AAG13898 standard; Protein; 322 AA.
XX
AC   AAG13898;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 13559.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
```

PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.

```
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             18.3%;  Score 364.5;  DB 21;  Length 322;
  Best Local Similarity   30.5%;  Pred. No. 5.3e-20;
  Matches  100;  Conservative   63;  Mismatches  132;  Indels   33;  Gaps   11;

Qy         53 YRGNIYNIPICLWLLDTYPYNPPICFVKPTSSMTIK-TGKHVDANGKIYLPYLHDWKHPR 111
              : |  ||||: :|||::|| :||  :| ||: | ||    ||  :| : |||| :| :|
Db          3 FHGVTYNIPVIIWLLESYPRHPPCVYVNPTADMIIKRPHAHVTPSGLVSLPYLQNWVYPS 62

Qy        112 SELLELIQIMIVIFGEEPPVFSRPTVSASYPPYTATGPPNT--SYMPGMPSGISAYPSGY 169
              | |::|:  :   |  :||::||       ||     ||    | :   ||   : |  :
Db         63 SNLVDLVSDLSAAFARDPPLYSR---RRPQPP--PPSPPTVYDSSLSRPPSADQSLPRPF 117

Qy        170 PPNPSGYPGCPYPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISE--DTIRASLISAVSD 227
              ||:| |          |          :  |          :   |::  : : : |:|
Db        118 PPSPYG--------GGVSRVQVQHVHHQQQSDDAAEVFKRNAINKMVEMVHSDLVS--MR 167

Qy        228 KLRWRMKEEMDGAQAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDE 287
              : |    ||:   ||    ||| |::|  |   |:|||    :|   ::: ::::    :
Db        168 RAREAEAEELLSLQA---GLKRREDELNIG---LKEMV----EEKETLEQQLQIISMNTD 217

Qy        288 ELSSALEKMENQSEN---NDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGV 344
              | | : : : :::|    |:|        | ||:|   | : |||| |: | :: : ||
Db        218 ILDSWVRENQGKTKNLVDLDVDNAFECGDTLSKQMLECTALDLAIEDAIYSLDKSFQDGV 277

Qy        345 IDLDVFLKHVRLLSRKQFQLRALMQKAR 372
              :  | :|::|||||||:||  ||   | |
Db        278 VPFDQYLRNVRLLSREQFFHRATGSKVR 305


RESULT 13
AAB93473
ID   AAB93473 standard; Protein; 341 AA.
XX
AC   AAB93473;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human protein sequence SEQ ID NO:12750.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 12750; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence   341 AA;

  Query Match             18.1%;  Score 361;  DB 22;  Length 341;
  Best Local Similarity   59.0%;  Pred. No. 1.1e-19;
  Matches   62;  Conservative   21;  Mismatches   22;  Indels    0;  Gaps    0;

Qy         29 LDSYVFNDGSSRELVNLTGTIPVRYRGNIYNIPICLWLLDTYPYNPPICFVKPTSSMTIK 88
              :|:||| | | ::|:| |||||| |:|| |||||  |:||::|: |||||:|||::| |
Db          1 MDTYVFKDSSQKDLLNFTGTIPVMYQGNTYNIPIRFWILDSHPFAPPICFLKPTANMGIL 60

Qy         89 TGKHVDANGKIYLPYLHDWKHPRSELLELIQIMIVIFGEEPPVFS 133
              |||||| |:|||||| :| ||:| :: ||: ||  | || |::|
Db         61 VGKHVDAQGRIYLPYLQNWSHPKSVIVGLIKEMIAKFQEELPMYS 105


RESULT 14
AAB53717
ID   AAB53717 standard; Protein; 146 AA.
XX
AC   AAB53717;
XX
DT   09-MAR-2001  (first entry)
XX
DE   Human colon cancer antigen protein sequence SEQ ID NO:1257.
XX
KW   Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW   identification; cytostatic; cardioactive; neuroprotective; vulnerary;
KW   immunomodulatory; muscular; gynaecological; gastrointestinal;
KW   nephrotropic; antiinfective; antibacterial; gene therapy; wound;
KW   neural disorder; immune system disorder; muscular disorder;
KW   reproductive disorder; gastrointestinal disorder; renal disorder;
KW   infectious disease; cardiovascular disorder.
XX
OS   Homo sapiens.
XX
```

PN WO200055351-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05883.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2000-587534/55.
DR N-PSDB; AAC98474.
XX
PT Colon cancer associated gene sequences, referred to as colon cancer
PT antigens, useful for the treatment, prevention, and diagnosis of colon
PT disorders such as colon cancer -
XX
PS Claim 11; Page 1845; 2104pp; English.
XX
CC AAC97991 to AAC98763 encode the human colon cancer associated proteins,
CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
CC vulnerary, nephrotropic, antiinfective and antibacterial activities, and
CC can be used in gene therapy. The colon cancer antigen polynucleotides,
CC proteins and antibodies to the proteins are useful for the prevention,
CC treatment and diagnosis of colon disorders, such as colon cancer. The
CC polynucleotides may be used in diagnostics and research, such as for
CC chromosome identification, and as hybridisation probes. The proteins
CC may also be used to prevent diseases such as neural disorders, immune
CC system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, wounds, renal disorders, infectious
CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
CC AAB54007 represent sequences used in the exemplification of the present
CC invention.
XX
SQ Sequence 146 AA;

Query Match 16.4%; Score 327; DB 21; Length 146;
Best Local Similarity 90.9%; Pred. No. 1.4e-17;
Matches 60; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

```
Qy         33 VFNDGSSRELVNLTGTIPVRYRGNIYNIPICLWLLDTYPYNPPICFVKPTSSMTIKTGKH 92
              |||||||||:||||||| |||| ||||||||||||||||||||||||||||||||||||
Db         69 VFNDGSSRELMNLTGTIPVPYRGNTYNIPICLWLLDTYPYNPPICFVKPTSSMTIKTGKH 128

Qy         93 VDANGK 98
              ||   |
Db        129 VDXPKK 134
```

RESULT 15
AAU23116
ID AAU23116 standard; Protein; 452 AA.
XX
AC AAU23116;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human enzyme polypeptide #202.
XX
KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;
KW nephrotropic; anticoagulant.
XX
OS Homo sapiens.
XX
PN WO200155301-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01239.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.

PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX

DR WPI; 2001-465566/50.
DR N-PSDB; AAS40986.
XX
PT Novel polypeptides and polynucleotides useful for diagnosing,
PT preventing, treating neural, immune system, muscular, reproductive,
PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous
PT diseases -
XX
PS Claim 11; SEQ ID No 1112; 1180pp; English.
XX
CC The present invention relates to the isolation of novel human enzyme
CC polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences
CC encoding them. The enzyme polypeptides of the invention may comprise the
CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
CC isomerases or ligases. The sequences of the invention are useful in the
CC diagnosis, treatment, prevention and/or prognosis of a wide range of
CC disorders including hyperproliferative disorders (e.g. cancer),
CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders
CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders
CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
CC blood-related disorders (e.g. haemophilia), reproductive disorders
CC (e.g. infertility) and infectious disorders (e.g. Influenza). The
CC polynucleotides of the invention can also be used in gene therapy.
CC AAU22915-AAU23814 represent the novel human enzyme polypeptides of the
CC invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 452 AA;

Query Match 15.9%; Score 318; DB 22; Length 452;
Best Local Similarity 45.1%; Pred. No. 3.3e-16;
Matches 60; Conservative 21; Mismatches 30; Indels 22; Gaps 1;

```
Qy          1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNI 60
              :  |||:||||| :  ||   :   |  :|:|                      || |||
Db         14 LXGKYKFRDLTVEELRNVNVFFPHFKYSMDTY--------------------GNTYNI 51

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQI 120
              ||  |:||::|: |||||:|||::| |  |||||| |:|||||| :| ||:| :: ||:
Db         52 PIRFWILDSHPFAPPICFLKPTANMGILVGKHVDAQGRIYLPYLQNWSHPKSVIVGLIKE 111

Qy        121 MIVIFGEEPPVFS 133
              ||  | || |::|
Db        112 MIAKFQEELPMYS 124
```

Search completed: June 3, 2003, 14:07:49
Job time : 53.5677 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: June 3, 2003, 14:06:00 ; Search time 13.0171 Seconds
(without alignments)
861.186 Million cell updates/sec

Title: US-09-804-690-2
Perfect score: 1996
Sequence: 1 MMSKYKYRDLTVRQTVNVIA.....FQLRALMQKARKTAGLSDLY 381

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1996	100.0	381	1	US-08-585-758A-2	Sequence 2, Appli
2	1996	100.0	381	1	US-08-977-818-2	Sequence 2, Appli
3	1996	100.0	381	2	US-08-670-274B-2	Sequence 2, Appli
4	1996	100.0	381	2	US-08-786-999-3	Sequence 3, Appli
5	1996	100.0	381	4	US-09-146-187-2	Sequence 2, Appli
6	1900.5	95.2	380	1	US-08-585-758A-4	Sequence 4, Appli
7	1900.5	95.2	380	1	US-08-977-818-4	Sequence 4, Appli
8	1900.5	95.2	380	2	US-08-670-274B-4	Sequence 4, Appli
9	1900.5	95.2	380	4	US-09-146-187-4	Sequence 4, Appli
10	1900.5	95.2	390	2	US-08-786-999-1	Sequence 1, Appli
11	1022.5	51.2	237	4	US-08-999-774A-8	Sequence 8, Appli
12	150	7.5	466	3	US-08-526-136-13	Sequence 13, Appl
13	141.5	7.1	905	2	US-08-574-959A-9	Sequence 9, Appli
14	141.5	7.1	905	4	US-09-357-014-9	Sequence 9, Appli
15	141.5	7.1	1135	2	US-08-574-959A-7	Sequence 7, Appli
16	141.5	7.1	1135	4	US-09-357-014-7	Sequence 7, Appli
17	122.5	6.1	503	3	US-08-526-136-2	Sequence 2, Appli
18	122.5	6.1	505	3	US-08-526-136-4	Sequence 4, Appli
19	121.5	6.1	214	1	US-08-217-327-4	Sequence 4, Appli
20	121.5	6.1	723	6	5200183-4	Patent No. 5200183
21	120	6.0	1162	2	US-08-728-323A-2	Sequence 2, Appli
22	120	6.0	1162	4	US-09-298-568-2	Sequence 2, Appli
23	119	6.0	380	2	US-09-026-587-4	Sequence 4, Appli
24	119	6.0	380	2	US-09-227-420-4	Sequence 4, Appli
25	118.5	5.9	543	4	US-09-535-008-63	Sequence 63, Appl
26	118.5	5.9	577	4	US-09-535-008-61	Sequence 61, Appl
27	118.5	5.9	1646	4	US-09-535-008-67	Sequence 67, Appl
28	118.5	5.9	1647	4	US-09-535-008-2	Sequence 2, Appli
29	118.5	5.9	1649	4	US-09-535-008-75	Sequence 75, Appl
30	118.5	5.9	1650	4	US-09-535-008-71	Sequence 71, Appl
31	118.5	5.9	1678	4	US-09-535-008-69	Sequence 69, Appl
32	118.5	5.9	1679	4	US-09-535-008-65	Sequence 65, Appl
33	118.5	5.9	1681	4	US-09-535-008-77	Sequence 77, Appl
34	118.5	5.9	1682	4	US-09-535-008-73	Sequence 73, Appl
35	117.5	5.9	262	3	US-08-946-914-14	Sequence 14, Appl
36	117.5	5.9	262	4	US-09-656-450-14	Sequence 14, Appl
37	117.5	5.9	1185	4	US-09-041-886-23	Sequence 23, Appl
38	116	5.8	338	1	US-08-218-686-2	Sequence 2, Appli
39	116	5.8	338	3	US-08-460-242-2	Sequence 2, Appli
40	116	5.8	809	5	PCT-US91-01726-3	Sequence 3, Appli
41	115	5.8	652	6	5202236-13	Patent No. 5202236
42	110.5	5.5	902	1	US-08-396-479B-6	Sequence 6, Appli
43	110.5	5.5	902	1	US-08-818-823-6	Sequence 6, Appli
44	110	5.5	1017	4	US-09-600-776-6	Sequence 6, Appli
45	108.5	5.4	531	4	US-09-134-001C-4920	Sequence 4920, Ap

ALIGNMENTS

```
RESULT 1
US-08-585-758A-2
; Sequence 2, Application US/08585758A
; Patent No. 5679523
;  GENERAL INFORMATION:
;    APPLICANT:  LI, Limin
;    APPLICANT:  COHEN, Stanley N.
;    TITLE OF INVENTION:  METHOD FOR CONCURRENT DISRUPTION OF
;    TITLE OF INVENTION:  EXPRESSION OF MULTIPLE ALLELES OF MAMMALIAN GENES
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  FOUR EMBARCADERO CENTER, SUITE 3400
;      CITY:  SAN FRANCISCO
;      STATE:  CA
;      COUNTRY:  US
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/585,758A
;      FILING DATE:  12-JAN-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rowland, Bertram I
;      REGISTRATION NUMBER:  20015
;      REFERENCE/DOCKET NUMBER:  A62783/BIR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-781-1989
;      TELEFAX:  415-398-3249
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  381 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-585-758A-2

  Query Match           100.0%;  Score 1996;  DB 1;  Length 381;
  Best Local Similarity  100.0%;  Pred. No. 2.8e-158;
  Matches  381;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNI 60

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQI 120
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQI 120

Qy        121 MIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCP 180
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 MIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCP 180

Qy        181 YPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGA 240
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 YPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGA 240

Qy        241 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 300
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 300

Qy        301 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 360
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 360

Qy        361 QFQLRALMQKARKTAGLSDLY 381
                 |||||||||||||||||||
Db        361 QFQLRALMQKARKTAGLSDLY 381


RESULT 2
US-08-977-818-2
; Sequence 2, Application US/08977818
; Patent No. 5807995
;  GENERAL INFORMATION:
;    APPLICANT:  LI, Limin
;    APPLICANT:  COHEN, Stanley N
;    TITLE OF INVENTION:  MAMMALIAN TUMOR SUSCEPTIBILITY GENES AND
;    TITLE OF INVENTION:  THEIR USES
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FISH AND RICHARDSON, P.C.
;      STREET:  2200 SAND HILL ROAD
;      CITY:  MENLO PARK
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94025
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/977,818
;      FILING DATE:  25-NOV-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/670,274
;      FILING DATE:  June 13, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SHERWOOD, Pamela J.
;      REGISTRATION NUMBER:  36,677
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-781-1989
;      TELEFAX:  415-398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  381 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-977-818-2

  Query Match             100.0%;  Score 1996;  DB 1;  Length 381;
  Best Local Similarity   100.0%;  Pred. No. 2.8e-158;
  Matches  381;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNI 60
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNI 60

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQI 120
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQI 120

Qy        121 MIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCP 180
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 MIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCP 180

Qy        181 YPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGA 240
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 YPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGA 240

Qy        241 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 300
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 300

Qy        301 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 360
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 360

Qy        361 QFQLRALMQKARKTAGLSDLY 381
                 |||||||||||||||||||
Db        361 QFQLRALMQKARKTAGLSDLY 381


RESULT 3
US-08-670-274B-2
; Sequence 2, Application US/08670274B
; Patent No. 5891668
;  GENERAL INFORMATION:
;    APPLICANT:  LI, Limin
;    APPLICANT:  COHEN, Stanley N
;    TITLE OF INVENTION:  MAMMALIAN TUMOR SUSCEPTIBILITY GENES AND
;    TITLE OF INVENTION:  THEIR USES
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FISH AND RICHARDSON, P.C.
;      STREET:  2200 SAND HILL ROAD
;      CITY:  MENLO PARK
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94025
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/670,274B
;      FILING DATE:  June 13, 1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SHERWOOD, Pamela J.
;      REGISTRATION NUMBER:  36,677
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-781-1989
;      TELEFAX:  415-398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  381 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-670-274B-2

  Query Match             100.0%;  Score 1996;  DB 2;  Length 381;
  Best Local Similarity   100.0%;  Pred. No. 2.8e-158;
  Matches  381;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNI 60

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQI 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQI 120

Qy        121 MIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCP 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 MIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCP 180

Qy        181 YPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 YPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGA 240

Qy        241 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 300

Qy        301 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 360

Qy        361 QFQLRALMQKARKTAGLSDLY 381
              |||||||||||||||||||||
Db        361 QFQLRALMQKARKTAGLSDLY 381
```

RESULT 4
US-08-786-999-3
; Sequence 3, Application US/08786999
; Patent No. 5892016
; GENERAL INFORMATION:
; APPLICANT: La Brie, Sam
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN TUMOR
; TITLE OF INVENTION: SUPPRESSOR
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/786,999
; FILING DATE: Filed Herewith
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0199 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 381 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank

DNA patented

; CLONE: 1330330
US-08-786-999-3

Query Match 100.0%; Score 1996; DB 2; Length 381;
Best Local Similarity 100.0%; Pred. No. 2.8e-158;
Matches 381; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNI 60

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQI 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQI 120

Qy        121 MIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCP 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 MIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCP 180

Qy        181 YPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 YPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGA 240

Qy        241 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 300

Qy        301 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 360

Qy        361 QFQLRALMQKARKTAGLSDLY 381
              |||||||||||||||||||||
Db        361 QFQLRALMQKARKTAGLSDLY 381
```

RESULT 5
US-09-146-187-2
; Sequence 2, Application US/09146187
; Patent No. 6248523
; GENERAL INFORMATION:
; APPLICANT: LI, Limin
; APPLICANT: COHEN, Stanley N
; TITLE OF INVENTION: MAMMALIAN TUMOR SUSCEPTIBILITY GENES AND
; TITLE OF INVENTION: THEIR USES
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH AND RICHARDSON, P.C.
; STREET: 2200 SAND HILL ROAD
; CITY: MENLO PARK
; STATE: CA
; COUNTRY: USA
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/146,187
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/670,274
; FILING DATE: June 13, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: SHERWOOD, Pamela J.
; REGISTRATION NUMBER: 36,677
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 2:

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  381 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-146-187-2

  Query Match             100.0%;  Score 1996;  DB 4;  Length 381;
  Best Local Similarity   100.0%;  Pred. No. 2.8e-158;
  Matches  381;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNI 60

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQI 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQI 120

Qy        121 MIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCP 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 MIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCP 180

Qy        181 YPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 YPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGA 240

Qy        241 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 300

Qy        301 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 360

Qy        361 QFQLRALMQKARKTAGLSDLY 381
              |||||||||||||||||||||
Db        361 QFQLRALMQKARKTAGLSDLY 381


RESULT 6
US-08-585-758A-4
; Sequence 4, Application US/08585758A
; Patent No. 5679523
;  GENERAL INFORMATION:
;    APPLICANT:  LI, Limin
;    APPLICANT:  COHEN, Stanley N.
;    TITLE OF INVENTION:  METHOD FOR CONCURRENT DISRUPTION OF
;    TITLE OF INVENTION:  EXPRESSION OF MULTIPLE ALLELES OF MAMMALIAN GENES
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  FOUR EMBARCADERO CENTER, SUITE 3400
;      CITY:  SAN FRANCISCO
;      STATE:  CA
;      COUNTRY:  US
;      ZIP:  94111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/585,758A
;      FILING DATE:  12-JAN-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rowland, Bertram I
;      REGISTRATION NUMBER:  20015
;      REFERENCE/DOCKET NUMBER:  A62783/BIR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-781-1989
;      TELEFAX:  415-398-3249
```

```
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  380 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-585-758A-4

  Query Match             95.2%;  Score 1900.5;  DB 1;  Length 380;
  Best Local Similarity   94.5%;  Pred. No. 2.5e-150;
  Matches  360;  Conservative   11;  Mismatches    9;  Indels    1;  Gaps    1;

Qy          1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNI 60
              |:|||||||||||:||||| :|||||||||||||||||||:||||||| |||| |||
Db          1 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 60

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQI 120
              ||||||||||||||||||||||||||||||||||||||||||:||||:|:|| |||:
Db         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120

Qy        121 MIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCP 180
              |||:||:|||||||| :||||||| ||||||||||||| ||| ||||||||||||||||
Db        121 MIVVFGDEPPVFSRP-ISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCP 179

Qy        181 YPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGA 240
              ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        180 YPPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRA 239

Qy        241 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        240 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 299

Qy        301 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        300 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 359

Qy        361 QFQLRALMQKARKTAGLSDLY 381
              |||||||||||||||||||||
Db        360 QFQLRALMQKARKTAGLSDLY 380


RESULT 7
US-08-977-818-4
; Sequence 4, Application US/08977818
; Patent No. 5807995
;  GENERAL INFORMATION:
;    APPLICANT:  LI, Limin
;    APPLICANT:  COHEN, Stanley N
;    TITLE OF INVENTION:  MAMMALIAN TUMOR SUSCEPTIBILITY GENES AND
;    TITLE OF INVENTION:  THEIR USES
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FISH AND RICHARDSON, P.C.
;      STREET:  2200 SAND HILL ROAD
;      CITY:  MENLO PARK
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94025
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/977,818
;      FILING DATE:  25-NOV-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/670,274
;      FILING DATE:  June 13, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SHERWOOD, Pamela J.
;      REGISTRATION NUMBER:  36,677
```

```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-781-1989
;      TELEFAX:  415-398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  380 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-977-818-4

  Query Match             95.2%;  Score 1900.5;  DB 1;  Length 380;
  Best Local Similarity   94.5%;  Pred. No. 2.5e-150;
  Matches  360;  Conservative   11;  Mismatches    9;  Indels    1;  Gaps    1;

Qy          1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVYRYRGNIYNI 60
              |:||||||||||||:||||| :|||||||||||||||||||||:|||||||| |||| |||
Db          1 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 60

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQI 120
              ||||||||||||||||||||||||||||||||||||||||||||:||||:|:|| |||:
Db         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120

Qy        121 MIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCP 180
              |||:||:|||||||| :|||||| |||||||||||| ||| ||||||||||||||
Db        121 MIVVFGDEPPVFSRP-ISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCP 179

Qy        181 YPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGA 240
              ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        180 YPPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRA 239

Qy        241 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        240 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 299

Qy        301 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        300 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 359

Qy        361 QFQLRALMQKARKTAGLSDLY 381
              |||||||||||||||||||||
Db        360 QFQLRALMQKARKTAGLSDLY 380


RESULT 8
US-08-670-274B-4
; Sequence 4, Application US/08670274B
; Patent No. 5891668
;  GENERAL INFORMATION:
;    APPLICANT:  LI, Limin
;    APPLICANT:  COHEN, Stanley N
;    TITLE OF INVENTION:  MAMMALIAN TUMOR SUSCEPTIBILITY GENES AND
;    TITLE OF INVENTION:  THEIR USES
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FISH AND RICHARDSON, P.C.
;      STREET:  2200 SAND HILL ROAD
;      CITY:  MENLO PARK
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94025
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/670,274B
;      FILING DATE:  June 13, 1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SHERWOOD, Pamela J.
;      REGISTRATION NUMBER:  36,677
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-781-1989
;      TELEFAX:  415-398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  380 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-670-274B-4

  Query Match             95.2%;  Score 1900.5;  DB 2;  Length 380;
  Best Local Similarity   94.5%;  Pred. No. 2.5e-150;
  Matches  360;  Conservative   11;  Mismatches    9;  Indels    1;  Gaps    1;

Qy          1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVYRYRGNIYNI 60
              |:||||||||||||:||||| :|||||||||||||||||||||:|||||||| |||| |||
Db          1 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 60

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQI 120
              ||||||||||||||||||||||||||||||||||||||||||||:||||:|:|| |||:
Db         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120

Qy        121 MIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCP 180
              |||:||:|||||||| :|||||| |||||||||||| ||| ||||||||||||||
Db        121 MIVVFGDEPPVFSRP-ISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCP 179

Qy        181 YPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGA 240
              ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        180 YPPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRA 239

Qy        241 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        240 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 299

Qy        301 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        300 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 359

Qy        361 QFQLRALMQKARKTAGLSDLY 381
              |||||||||||||||||||||
Db        360 QFQLRALMQKARKTAGLSDLY 380


RESULT 9
US-09-146-187-4
; Sequence 4, Application US/09146187
; Patent No. 6248523
;  GENERAL INFORMATION:
;    APPLICANT:  LI, Limin
;    APPLICANT:  COHEN, Stanley N
;    TITLE OF INVENTION:  MAMMALIAN TUMOR SUSCEPTIBILITY GENES AND
;    TITLE OF INVENTION:  THEIR USES
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FISH AND RICHARDSON, P.C.
;      STREET:  2200 SAND HILL ROAD
;      CITY:  MENLO PARK
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94025
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/146,187
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  08/670,274
;      FILING DATE:  June 13, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SHERWOOD, Pamela J.
;      REGISTRATION NUMBER:  36,677
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-781-1989
;      TELEFAX:  415-398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  380 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-146-187-4

  Query Match             95.2%;  Score 1900.5;  DB 4;  Length 380;
  Best Local Similarity   94.5%;  Pred. No. 2.5e-150;
  Matches  360;  Conservative   11;  Mismatches    9;  Indels    1;  Gaps    1;

Qy          1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNI 60
              |:||||||||||:|||| :|||||||||||||||||||:|||||||| |||| |||
Db          1 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 60

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQI 120
              ||||||||||||||||||||||||||||||||||||||||:||||:|:|| |||:
Db         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120

Qy        121 MIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCP 180
              |||:||:|||||||| :|||||| |||||||||||| ||| ||||||||||||||
Db        121 MIVVFGDEPPVFSRP-ISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCP 179

Qy        181 YPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGA 240
              ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        180 YPPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRA 239

Qy        241 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        240 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 299

Qy        301 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        300 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 359

Qy        361 QFQLRALMQKARKTAGLSDLY 381
              |||||||||||||||||||||
Db        360 QFQLRALMQKARKTAGLSDLY 380


RESULT 10
US-08-786-999-1
; Sequence 1, Application US/08786999
; Patent No. 5892016
;  GENERAL INFORMATION:
;    APPLICANT:  La Brie, Sam
;    APPLICANT:  Goli, Surya K.
;    TITLE OF INVENTION:  NOVEL HUMAN TUMOR
;    TITLE OF INVENTION:  SUPPRESSOR
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US/08/786,999
;      FILING DATE:  Filed Herewith
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0199 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  390 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  Colnnot01
;      CLONE:  609476
US-08-786-999-1

  Query Match             95.2%;  Score 1900.5;  DB 2;  Length 390;
  Best Local Similarity   94.5%;  Pred. No. 2.6e-150;
  Matches  360;  Conservative   11;  Mismatches    9;  Indels    1;  Gaps    1;

Qy          1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNI 60
              |:||||||||||:|||| :|||||||||||||||||||:|||||||| |||| |||
Db         11 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 70

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQI 120
              ||||||||||||||||||||||||||||||||||||||||:||||:|:|| |||:
Db         71 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 130

Qy        121 MIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCP 180
              |||:||:|||||||| :|||||| |||||||||||| ||| ||||||||||||||
Db        131 MIVVFGDEPPVFSRP-ISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCP 189

Qy        181 YPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGA 240
              ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        190 YPPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRA 249

Qy        241 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        250 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 309

Qy        301 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        310 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 369

Qy        361 QFQLRALMQKARKTAGLSDLY 381
              |||||||||||||||||||||
Db        370 QFQLRALMQKARKTAGLSDLY 390


RESULT 11
US-08-999-774A-8
; Sequence 8, Application US/08999774A
; Patent No. 6274312
;  GENERAL INFORMATION:
;    APPLICANT:  Gish, Kurt C.
;    APPLICANT:  Seghezzi, Wolfgang
;    APPLICANT:  Shanahan, Frances
;    APPLICANT:  Lees; Emma M.
;    APPLICANT:  McClanahan, Terrill K.
;    TITLE OF INVENTION:  Intracellular Regulatory Molecules;
;    TITLE OF INVENTION:  Related Reagents
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DNAX Research Institute
;      STREET:  901 California Avenue
```

```
;       CITY:  Palo Alto
;       STATE:  California
;       COUNTRY:  USA
;       ZIP:  94304-1104
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/999,774A
;       FILING DATE:  10-DEC-1997
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 60/032,818
;       FILING DATE:  11-DEC-1996
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Ching, Edwin P.
;       REGISTRATION NUMBER:  34,090
;       REFERENCE/DOCKET NUMBER:  DX0646
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (650)852-9196
;       TELEFAX:  (650)496-1200
;   INFORMATION FOR SEQ ID NO:  8:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  237 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-999-774A-8

  Query Match             51.2%;  Score 1022.5;  DB 4;  Length 237;
  Best Local Similarity   88.3%;  Pred. No. 1.5e-77;
  Matches  182;  Conservative   12;  Mismatches   11;  Indels    1;  Gaps    1;

Qy          1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNI 60
              |:||||||||||:|||| :|||||||||||||||||||:|||||| |||| |||
Db         11 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 70

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQI 120
              ||||||||||||||||||||||||||||||||||||||||||:||||:|:|| |||: '
Db         71 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 130

Qy        121 MIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCP 180
              |||:||:|||||||| :|||||| |||||||||||||| ||| ||||||||||||||||
Db        131 MIVVFGDEPPVFSRP-ISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCP 189

Qy        181 YPPAGPYPATTSSQYPSQPPVTTVGP 206
              ||| ||||||||||||||||  : |
Db        190 YPPGGPYPATTSSQYPSQPPVPLLVP 215


RESULT 12
US-08-526-136-13
; Sequence 13, Application US/08526136
; Patent No. 6107089
;  GENERAL INFORMATION:
;    APPLICANT:  Towle, Christine A. et al.
;    TITLE OF INVENTION:  ANNEXIN XI
;    NUMBER OF SEQUENCES:  36
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      COMPUTER:  IBM PS/2 Model 50Z or 55SX
;      OPERATING SYSTEM:  IBM P.C. DOS (Version 3.30)
;      SOFTWARE:  WordPerfect (Version 5.0)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/526,136
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/214,036
;      FILING DATE:
;      APPLICATION NUMBER:  07/837,775
;      FILING DATE:  February 13, 1992
;      APPLICATION NUMBER:  07/764,465
;      FILING DATE:  September 23, 1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Paul T.
;      REGISTRATION NUMBER:  30,162
;      REFERENCE/DOCKET NUMBER:  00786/099001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 542-5070
;      TELEFAX:  (617) 542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  466
;      TYPE:  amino acid
;      STRANDEDNESS:  N/A
;      TOPOLOGY:  N/A
US-08-526-136-13

  Query Match             7.5%;  Score 150;  DB 3;  Length 466;
  Best Local Similarity   21.3%;  Pred. No. 0.00017;
  Matches   83;  Conservative   36;  Mismatches   92;  Indels  178;  Gaps   16;

Qy        140 SYPPYTATG-PPNTSYMPG------MPSGISAYPSGYP---------------------- 170
              ||| |  || ||   | |        |||   ||||:|
Db          2 SYPGYPPTGYPPFPGYPPAGQESSFPPSGQYPYPSGFPPMGGGAYPQVPSSGYPGAGGYP 61

Qy        171 -----PNPSGYPGCPYPPAGP-YPATTSSQ------------YPSQPPVTTVG------ 205
                  | | |||| | |   | ||     |                || |||  : |
Db         62 APGGYPAPGGYPGAPQPGGAPSYPGVPPGQGFGVPPGGAGFSGYP-QPPSQSYGGGPAQV 120

Qy        206 ----------------------PSRDGTISEDT--------------------------- 216
                                     ||:  |::: |
Db        121 PLPGGFPGGQMPSQYPGGQPTYPSQPATVTQVTQGTIRPAANFDAIRDAEILRKAMKGFG 180

Qy        217 --------------------IRASLISAVSDKLRWRMKEEMDGAQAELNA---LKRTEED 253
                                 |:|:  ::    |   :| |: |   ||     : | |
Db        181 TDEQAIVDVVANRSNDQRQKIKAAFKTSYGKDLIKDLKSELSGNMEELILALFMPPTYYD 240

Qy        254 ---LKKGHQ--------KLEEMVTRLDQEVAEV---DKNIELLKKDEELSSALEKM 296
                |:|  |         :| : || :||: |:      :   :| |    : |   |::
Db        241 AWSLRKAMQGAGTQERVLIEILCTRTNQEIREIVRCYQSEFGRDLEKDIRSDTSGHFERL 300

Qy        297 ----------ENQSENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVID 346
                         |||| |: :                |:|:|    ::  ||    | :
Db        301 LVSMCQGNRDENQSINHQM----------------AQEDA--QRLYQAGE----GRLG 336

Qy        347 LDVFLKHVRLLSRKQFQLRALMQKARKTA 375
              |    :: | :|   |||| |:   : |
Db        337 TDESCFNMILATRSFPQLRATMEAYSRMA 365


RESULT 13
US-08-574-959A-9
; Sequence 9, Application US/08574959A
; Patent No. 5962224
;  GENERAL INFORMATION:
;    APPLICANT:  Jaekyoon Shin, Insil Joung, Ratna K. Vadlamudi
;    APPLICANT:  and Jack L. Strominger
;    TITLE OF INVENTION:  p62 POLYPEPTIDES, RELATED POLYPEPTIDES
;    TITLE OF INVENTION:  AND USES THEREFOR
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 State Street, Suite 510
;      CITY:  Boston
```

```
;       STATE:  Massachusetts
;       COUNTRY:  USA
;       ZIP:  02109-1875
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/574,959A
;       FILING DATE:  19-DEC-95
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Mandragouras, Amy E.
;       REGISTRATION NUMBER:  36,207
;       REFERENCE/DOCKET NUMBER:  DFN-008
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617)227-7400
;       TELEFAX:  (617)227-5941
;   INFORMATION FOR SEQ ID NO:  9:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  905 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-574-959A-9

  Query Match             7.1%;  Score 141.5;  DB 2;  Length 905;
  Best Local Similarity   26.0%;  Pred. No. 0.0022;
  Matches   50;  Conservative   32;  Mismatches   83;  Indels   27;  Gaps    7;

Qy        122 IVIFGEE-PPVFSRPTVSASYPPYTATGPPNTSYMPGMPS-----GISAYPSGYPPNPSG 175
              :|| | ||:  |  |: ||  ||||  |  | :|:      : | |   || |
Db        554 VVIVPEGLPPLPPPPPSGATPPPIAPTGPPTAS--PPVPAKEEPEELPAAPGPLPPPP-- 609

Qy        176 YPGCPYPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKE 235
                  ||  | |       |   | | |      : ||    :: |  ||:     :|
Db        610 ------PPPPPVPGPVXLPPPQLVPEGTPGGGGPPALEEDLTVININS--SDE-----EE 656

Qy        236 EMDGAQAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEK 295
              | :| : |    :  ||: ::  :: ||     ::|  : ::  |  ::::||     |:
Db        657 EEEGEEEE----EEEEEEEEEEEEEEEEEDFEEEEDEEEYFEEEEEEEEEFEEEFEE 712

Qy        296 MENQSENNDIDE 307
              | : |  : :|
Db        713 EEGELEEEEEEE 724


RESULT 14
US-09-357-014-9
; Sequence 9, Application US/09357014
; Patent No. 6291645
;   GENERAL INFORMATION:
;       APPLICANT: Jaekyoon Shin, Insil Joung, Ratna K. Vadlamudi
;                  and Jack L. Strominger
;       TITLE OF INVENTION: p62 POLYPEPTIDES, RELATED POLYPEPTIDES
;                           AND USES THEREFOR
;       NUMBER OF SEQUENCES: 22
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: LAHIVE & COCKFIELD
;            STREET: 60 State Street, Suite 510
;            CITY: Boston
;            STATE: Massachusetts
;            COUNTRY: USA
;            ZIP: 02109-1875
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.25
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/357,014
;            FILING DATE: 19-Jul-1999
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/574,959
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Mandragouras, Amy E.
;            REGISTRATION NUMBER: 36,207
;            REFERENCE/DOCKET NUMBER: DFN-008
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (617)227-7400
;            TELEFAX: (617)227-5941
;   INFORMATION FOR SEQ ID NO: 9:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 905 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-09-357-014-9

  Query Match             7.1%;  Score 141.5;  DB 4;  Length 905;
  Best Local Similarity   26.0%;  Pred. No. 0.0022;
  Matches   50;  Conservative   32;  Mismatches   83;  Indels   27;  Gaps    7;

Qy        122 IVIFGEE-PPVFSRPTVSASYPPYTATGPPNTSYMPGMPS-----GISAYPSGYPPNPSG 175
              :|| | ||:  |  |: ||  ||||  |  | :|:      : | |   || |
Db        554 VVIVPEGLPPLPPPPPSGATPPPIAPTGPPTAS--PPVPAKEEPEELPAAPGPLPPPP-- 609

Qy        176 YPGCPYPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKE 235
                  ||  | |       |   | | |      : ||    :: |  ||:     :|
Db        610 ------PPPPPVPGPVXLPPPQLVPEGTPGGGGPPALEEDLTVININS--SDE-----EE 656

Qy        236 EMDGAQAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEK 295
              | :| : |    :  ||: ::  :: ||     ::|  : ::  |  ::::||     |:
Db        657 EEEGEEEE----EEEEEEEEEEEEEEEEEDFEEEEDEEEYFEEEEEEEEEFEEEFEE 712

Qy        296 MENQSENNDIDE 307
              | : |  : :|
Db        713 EEGELEEEEEEE 724


RESULT 15
US-08-574-959A-7
; Sequence 7, Application US/08574959A
; Patent No. 5962224
;  GENERAL INFORMATION:
;    APPLICANT:  Jaekyoon Shin, Insil Joung, Ratna K. Vadlamudi
;    APPLICANT:  and Jack L. Strominger
;    TITLE OF INVENTION:  p62 POLYPEPTIDES, RELATED POLYPEPTIDES
;    TITLE OF INVENTION:  AND USES THEREFOR
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 State Street, Suite 510
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/574,959A
;      FILING DATE:  19-DEC-95
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mandragouras, Amy E.
;      REGISTRATION NUMBER:  36,207
;      REFERENCE/DOCKET NUMBER:  DFN-008
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
```

; LENGTH: 1135 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-574-959A-7

Query Match 7.1%; Score 141.5; DB 2; Length 1135;
Best Local Similarity 26.0%; Pred. No. 0.0031;
Matches 50; Conservative 32; Mismatches 83; Indels 27; Gaps 7;

```
Qy        122 IVIFGEE-PPVFSRPTVSASYPPYTATGPPNTSYMPGMPS-----GISAYPSGYPPNPSG 175
              :||  |  ||:   |   |: ||   ||||  |  | :|:      : | |   || |
Db        784 VVIVPEGLPPLPPPPPSGATPPPIAPTGPPTAS--PPVPAKEEPEELPAAPGPLPPPP-- 839

Qy        176 YPGCPYPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKE 235
                  ||  | |       |  |  | |      : ||    :: |  ||:     :|
Db        840 ------PPPPPVPGPVXLPPPQLVPEGTPGGGGPPALEEDLTVININS--SDE-----EE 886

Qy        236 EMDGAQAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEK 295
              | :| : |    :  ||: ::  :: ||     ::|  : ::  |  ::::||     |:
Db        887 EEEGEEEE----EEEEEEEEEEEEEEEEEDFEEEEDEEEYFEEEEEEEEFEEEFEE 942

Qy        296 MENQSENNDIDE 307
              | : |  : :|
Db        943 EEGELEEEEEE 954
```

Search completed: June 3, 2003, 14:08:56
Job time : 15.0171 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: June 3, 2003, 14:09:05 ; Search time 30.5401 Seconds
(without alignments)
1262.807 Million cell updates/sec

Title: US-09-804-690-2
Perfect score: 1996
Sequence: 1 MMSKYKYRDLTVRQTVNVIA.....FQLRALMQKARKTAGLSDLY 381

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 383519 seqs, 101223694 residues

Total number of hits satisfying chosen parameters: 383519

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1996	100.0	381	10	US-09-804-690-2	Sequence 2, Appli
2	1900.5	95.2	380	10	US-09-804-690-4	Sequence 4, Appli
3	327	16.4	146	9	US-09-925-299-1257	Sequence 1257, Ap
4	327	16.4	146	10	US-09-925-299-1257	Sequence 1257, Ap
5	190.5	9.5	296	10	US-09-801-368-386	Sequence 386, App
6	152	7.6	485	10	US-09-925-300-1664	Sequence 1664, Ap
7	146	7.3	148	9	US-10-012-542-453	Sequence 453, App
8	126.5	6.3	416	9	US-10-043-487-282	Sequence 282, App
9	124.5	6.2	371	9	US-09-284-320-2	Sequence 2, Appli
10	124.5	6.2	371	9	US-10-197-666A-4	Sequence 4, Appli
11	121	6.1	887	10	US-09-771-161A-192	Sequence 192, App
12	121	6.1	887	10	US-09-771-161A-193	Sequence 193, App
13	119	6.0	380	10	US-09-823-240-10	Sequence 10, Appl
14	119	6.0	635	9	US-10-072-036-125	Sequence 125, App
15	118.5	5.9	92	9	US-10-083-357-1277	Sequence 1277, Ap
16	118.5	5.9	559	10	US-09-858-155A-2	Sequence 2, Appli
17	118.5	5.9	1647	9	US-09-824-574-4	Sequence 4, Appli
18	117.5	5.9	262	9	US-10-235-674-14	Sequence 14, Appl
19	117.5	5.9	262	10	US-09-263-689-14	Sequence 14, Appl
20	116	5.8	756	10	US-09-771-161A-101	Sequence 101, App
21	115.5	5.8	284	9	US-10-063-547-62	Sequence 62, Appl
22	115.5	5.8	284	9	US-10-174-590-236	Sequence 236, App
23	115.5	5.8	284	9	US-10-176-758-236	Sequence 236, App
24	115.5	5.8	284	9	US-10-063-616-62	Sequence 62, Appl
25	115.5	5.8	284	9	US-10-175-737-236	Sequence 236, App
26	115.5	5.8	284	9	US-10-063-502-62	Sequence 62, Appl
27	115.5	5.8	284	9	US-10-173-706-236	Sequence 236, App
28	115.5	5.8	284	9	US-10-175-738-236	Sequence 236, App
29	115.5	5.8	284	9	US-10-175-752-236	Sequence 236, App
30	115.5	5.8	284	9	US-10-176-482-236	Sequence 236, App
31	115.5	5.8	284	9	US-10-176-757-236	Sequence 236, App
32	115.5	5.8	284	9	US-10-176-913-236	Sequence 236, App
33	115.5	5.8	284	9	US-10-180-552-236	Sequence 236, App
34	115.5	5.8	284	9	US-10-180-557-236	Sequence 236, App
35	115.5	5.8	284	9	US-10-173-700-236	Sequence 236, App
36	115.5	5.8	284	9	US-10-174-572-236	Sequence 236, App
37	115.5	5.8	284	9	US-10-174-579-236	Sequence 236, App
38	115.5	5.8	284	9	US-10-174-582-236	Sequence 236, App
39	115.5	5.8	284	9	US-10-174-588-236	Sequence 236, App
40	115.5	5.8	284	9	US-10-175-739-236	Sequence 236, App
41	115.5	5.8	284	9	US-10-175-740-236	Sequence 236, App
42	115.5	5.8	284	9	US-10-175-743-236	Sequence 236, App
43	115.5	5.8	284	9	US-10-176-488-236	Sequence 236, App
44	115.5	5.8	284	9	US-10-176-492-236	Sequence 236, App
45	115.5	5.8	284	9	US-10-176-747-236	Sequence 236, App

ALIGNMENTS

```
RESULT 1
US-09-804-690-2
; Sequence 2, Application US/09804690
; Patent No. US20020034743A1
;  GENERAL INFORMATION:
;       APPLICANT: LI, Limin
;                  COHEN, Stanley N
;       TITLE OF INVENTION: MAMMALIAN TUMOR SUSCEPTIBILITY GENES AND
;                           THEIR USES
;       NUMBER OF SEQUENCES: 20
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: FISH AND RICHARDSON, P.C.
;            STREET: 2200 SAND HILL ROAD
;            CITY: MENLO PARK
;            STATE: CA
;            COUNTRY: USA
;            ZIP: 94025
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/804,690
;            FILING DATE: 12-Mar-2001
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 09/146,187
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: SHERWOOD, Pamela J.
;            REGISTRATION NUMBER: 36,677
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 415-781-1989
;            TELEFAX: 415-398-3249
;            TELEX: 910 277299
;  INFORMATION FOR SEQ ID NO: 2:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 381 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
```

```
;       SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-804-690-2

  Query Match             100.0%;  Score 1996;  DB 10;  Length 381;
  Best Local Similarity   100.0%;  Pred. No. 7e-130;
  Matches  381;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNI 60

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQI 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQI 120

Qy        121 MIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCP 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 MIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCP 180

Qy        181 YPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 YPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGA 240

Qy        241 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 300

Qy        301 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 360

Qy        361 QFQLRALMQKARKTAGLSDLY 381
              |||||||||||||||||||||
Db        361 QFQLRALMQKARKTAGLSDLY 381


RESULT 2
US-09-804-690-4
; Sequence 4, Application US/09804690
; Patent No. US20020034743A1
;   GENERAL INFORMATION:
;        APPLICANT: LI, Limin
;                   COHEN, Stanley N
;        TITLE OF INVENTION: MAMMALIAN TUMOR SUSCEPTIBILITY GENES AND
;                            THEIR USES
;        NUMBER OF SEQUENCES: 20
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: FISH AND RICHARDSON, P.C.
;             STREET: 2200 SAND HILL ROAD
;             CITY: MENLO PARK
;             STATE: CA
;             COUNTRY: USA
;             ZIP: 94025
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/804,690
;             FILING DATE: 12-Mar-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 09/146,187
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: SHERWOOD, Pamela J.
;             REGISTRATION NUMBER: 36,677
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 415-781-1989
;             TELEFAX: 415-398-3249
;             TELEX: 910 277299
;   INFORMATION FOR SEQ ID NO: 4:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 380 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-804-690-4

  Query Match             95.2%;  Score 1900.5;  DB 10;  Length 380;
  Best Local Similarity   94.5%;  Pred. No. 2.6e-123;
  Matches  360;  Conservative   11;  Mismatches    9;  Indels    1;  Gaps    1;

Qy          1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNI 60
              |:||||||||||:||||| :|||||||||||||||||||:|||||||| |||| |||
Db          1 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 60

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQI 120
              ||||||||||||||||||||||||||||||||||||||||:||||:|:|| |||:
Db         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 120

Qy        121 MIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCP 180
              |||:||:|||||||| :||||||| |||||||||||||| ||| |||||||||||||||
Db        121 MIVVFGDEPPVFSRP-ISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCP 179

Qy        181 YPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGA 240
              ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        180 YPPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRA 239

Qy        241 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        240 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 299

Qy        301 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        300 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 359

Qy        361 QFQLRALMQKARKTAGLSDLY 381
              |||||||||||||||||||||
Db        360 QFQLRALMQKARKTAGLSDLY 380


RESULT 3
US-09-925-299-1257
; Sequence 1257, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA102
;  CURRENT APPLICATION NUMBER: US/09/925,299
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05883
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 1556
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1257
;  LENGTH: 146
;  TYPE: PRT
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: SITE
;  LOCATION: (2)
;  OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;  NAME/KEY: SITE
;  LOCATION: (6)
;  OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;  NAME/KEY: SITE
;  LOCATION: (12)
;  OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;  NAME/KEY: SITE
;  LOCATION: (131)
```

```
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (135)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (138)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (145)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-1257

  Query Match             16.4%;  Score 327;  DB 9;  Length 146;
  Best Local Similarity   90.9%;  Pred. No. 1.7e-15;
  Matches   60;  Conservative    1;  Mismatches    5;  Indels    0;  Gaps    0;

Qy         33 VFNDGSSRELVNLTGTIPVRYRGNIYNIPICLWLLDTYPYNPPICFVKPTSSMTIKTGKH 92
              ||||||||||:|||||||| |||| ||||||||||||||||||||||||||||||||||||
Db         69 VFNDGSSRELMNLTGTIPVPYRGNTYNIPICLWLLDTYPYNPPICFVKPTSSMTIKTGKH 128

Qy         93 VDANGK 98
              ||   |
Db        129 VDXPKK 134


RESULT 4
US-09-925-299-1257
; Sequence 1257, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA102
;  CURRENT APPLICATION NUMBER: US/09/925,299
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05883
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 1556
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1257
;   LENGTH: 146
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (2)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (6)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (12)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (131)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (135)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (138)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (145)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-1257

  Query Match             16.4%;  Score 327;  DB 10;  Length 146;
  Best Local Similarity   90.9%;  Pred. No. 1.7e-15;
  Matches   60;  Conservative    1;  Mismatches    5;  Indels    0;  Gaps    0;

Qy         33 VFNDGSSRELVNLTGTIPVRYRGNIYNIPICLWLLDTYPYNPPICFVKPTSSMTIKTGKH 92
              ||||||||||:|||||||| |||| ||||||||||||||||||||||||||||||||||||
Db         69 VFNDGSSRELMNLTGTIPVPYRGNTYNIPICLWLLDTYPYNPPICFVKPTSSMTIKTGKH 128

Qy         93 VDANGK 98
              ||   |
Db        129 VDXPKK 134


RESULT 5
US-09-801-368-386
; Sequence 386, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
;  APPLICANT: Busby, Robert
;  APPLICANT:  Cali, Brian
;  APPLICANT:  Hecht, Peter
;  APPLICANT:  Holtzman, Doug
;  APPLICANT:  Madden, Kevin
;  APPLICANT:  Maxon, Mary
;  APPLICANT:  Milne, Todd
;  APPLICANT:  No. US20020128250Alman, Thea
;  APPLICANT:  Royer, John
;  APPLICANT:  Salama, Sofie
;  APPLICANT:  Sherman, Amir
;  APPLICANT:  Silva, Jeff
;  APPLICANT:  Summers, Eric
;  TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
;  FILE REFERENCE: 109272.147
;  CURRENT APPLICATION NUMBER: US/09/801,368
;  CURRENT FILING DATE:  2001-03-07
;  PRIOR APPLICATION NUMBER: US 09/487,558
;  PRIOR FILING DATE: 2000-01-19
;  PRIOR APPLICATION NUMBER: US 60/160,587
;  PRIOR FILING DATE: 1999-10-20
;  NUMBER OF SEQ ID NOS: 440
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 386
;   LENGTH: 296
;   TYPE: PRT
;   ORGANISM: Saccharomyces cerevisiae
US-09-801-368-386

  Query Match              9.5%;  Score 190.5;  DB 10;  Length 296;
  Best Local Similarity   23.6%;  Pred. No. 1.1e-05;
  Matches   71;  Conservative   54;  Mismatches  125;  Indels   51;  Gaps   13;

Qy          7 YRD--LTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNI-YNIPIC 63
              | |   |   :: ::  :  |:|    :  :||: : |::: |||     |:  ::|||:
Db         25 YNDGRTTFHDSLALLDNFHSLRPRTRVFTHSDGTPQLLLSIYGTISTGEDGSSPHSIPVI 84

Qy         64 LWLLDTYPYNPPIC------FVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLEL 117
              :|:   ||  ||        |   | | ::   :::|:|| | || || | |      |: :
Db         85 MWVPSMYPVKPPFISINLENFDMNTISSSLPIQEYIDSNGWIALPILHCWDPAAMNLIMV 144

Qy        118 IQIMIVIFGEEPPVFSRPTVSASYPPYTAT---GPPNTSYMPGMPSGISAYPSGYPPNPS 174
              :| :: :   |||    |    | ||   |      ||  :|  |      |   || |
Db        145 VQELMSLL-HEPPQDQAP----SLPPKPNTQLQQEQNTPPLPPKPKS----PHLKPPLP- 194

Qy        175 GYPGCPYPPAGPYPATTSSQYPSQPPVTTVGPSRD-------GTISEDTIRASLISAVSD 227
                    ||  | ||: :          | : |:          |:  :  |   :  |:|
Db        195 -------PPPPQPASNALDLMDMDN-TDISPTNHHEMLQNLQTVVNELYRED-VDYVAD 245

Qy        228 KLRWR---MKEEMDGAQAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKK 284
              |:  |   |:|          :: |  | :      |  :   ::| :  ::  |::| :
Db        246 KILTRQTVMQE----------SIARFHEIIAIDKNHLRAVEQAIEQTMHSLNAQIDVLNR 295

Qy        285 K 285
              |
Db        296 K 296


RESULT 6
```

```
US-09-925-300-1664
; Sequence 1664, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
;  APPLICANT: Craig Rosen,
;  APPLICANT:  Steve Ruben
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA101
;  CURRENT APPLICATION NUMBER: US/09/925,300
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05988
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 1890
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1664
;   LENGTH: 485
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-925-300-1664

  Query Match             7.6%;  Score 152;  DB 10;  Length 485;
  Best Local Similarity   21.4%;  Pred. No. 0.0092;
  Matches   84;  Conservative   36;  Mismatches   94;  Indels  178;  Gaps   16;

Qy        137 VSASYPPYTATG-PPNTSYMPG------MPSGISAYPSGYP------------------- 170
              |  ||| |  || ||   | |        |||   ||||:|
Db         18 VRMSYPGYPPTGYPPFPGYPPAGQESSFPPSGQYPYPSGFPPPMGGGAYPQVPSSGYPGAG 77

Qy        171 --------PNPSGYPGCPYPPAGP-YPATTSSQ-------------YPSQPPVTTVG--- 205
                      | | |||| | |    | ||      |                || |||  : |
Db         78 GYPAPGGYPAPGGYPGAPQPGGAPSYPGVPPGQGFGVPPGGAGFSGYP-QPPSQSYGGGP 136

Qy        206 ------------------------PSRDGTISEDT------------------------ 216
                                      ||:  |::: |
Db        137 AQVPLPGGFPGGQMPSQYPGGQPTYPSQPATVTQVTQGTIRPAANFDAIRDAEILRKAMK 196

Qy        217 ----------------------IRASLISAVSDKLRWRMKEEMDGAQAELNA---LKRT 250
                                   |:|:  ::    |   :| |: |   ||      :  |
Db        197 GFGTDEQAIVDVVANRSNDQRQKIKAAFKTSYGKDLIKDLKSELSGNMEELILALFMPPT 256

Qy        251 EED---LKKGHQ--------KLEEMVTRLDQEVAEV------DKNIELLKKKDEELSSAL 293
                |   |:|  |         :| : || :||: |:       :   :| |    : |
Db        257 YYDAWSLRKAMQGAGTQERVLIEILCTRTNQEIREIVRCYQSEFGRDLEKDIRSDTSGHF 316

Qy        294 EKM----------ENQSENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRG 343
              |::          |||| |: :                  |:|:|    ::  ||    |
Db        317 ERLLVSMCQGNRDENQSINHQM-----------------AQEDA--QRLYQAGE----G 352

Qy        344 VIDLLDVFLKHVRLLSRKQFQLRALMQKARKTA 375
                :  |    :: | :|   |||| |:   : |
Db        353 RLGTDESCFNMILATRSFPQLRATMEAYSRMA 384


RESULT 7
US-10-012-542-453
; Sequence 453, Application US/10012542
; Publication No. US20030044851A1
; GENERAL INFORMATION:
;  APPLICANT: Ruben et al.
;  TITLE OF INVENTION: 94 Human Secreted Proteins
;  FILE REFERENCE: PZ029P1
;  CURRENT APPLICATION NUMBER: US/10/012,542
;  CURRENT FILING DATE:  2001-12-12
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/461,325
;  PRIOR FILING DATE: EARLIER FILING DATE: 1999-12-14
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,507
;  PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,508
;  PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,509
;  PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,510
;  PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/090,112
;  PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-22
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/090,113
;  PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-22
;  NUMBER OF SEQ ID NOS: 532
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 453
;   LENGTH: 148
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-012-542-453

  Query Match             7.3%;  Score 146;  DB 9;  Length 148;
  Best Local Similarity   27.9%;  Pred. No. 0.0051;
  Matches   38;  Conservative   15;  Mismatches   33;  Indels   50;  Gaps    6;

Qy        135 PTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNP---SGYPGC----PYPPAGPY 187
              | ::  :||     |      || | |  |:| | ||:|    |||||    |||| |
Db         30 PPINPPFPPGPCPPP------PGAPHGNPAFPPGGPPHPVPQPGYPGCQPLGPYPPPYPP 83

Qy        188 PATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGAQAELNAL 247
              ||       |  |||  : |   |             : : |
Db         84 PA------PGIPPVNPLAPGMVGP----------AVIVD-------------------- 106

Qy        248 KRTEEDLKKGHQKLEE 263
              |: :: :|| |:|: :
Db        107 KKMQKKMKKAHKKMHK 122


RESULT 8
US-10-043-487-282
; Sequence 282, Application US/10043487
; Publication No. US20030055220A1
; GENERAL INFORMATION:
;  APPLICANT: HYBRIGENICS
;  APPLICANT:  Pierre, LEGRAIN
;  TITLE OF INVENTION: Protein-protein interactions between Shigella Flexneri polypep
;  TITLE OF INVENTION:  mammalian polypeptides
;  FILE REFERENCE: B4778A
;  CURRENT APPLICATION NUMBER: US/10/043,487
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/261,130
;  PRIOR FILING DATE: 2001-01-12
;  NUMBER OF SEQ ID NOS: 561
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 282
;   LENGTH: 416
;   TYPE: PRT
;   ORGANISM: Shigella Flexneri
US-10-043-487-282

  Query Match             6.3%;  Score 126.5;  DB 9;  Length 416;
  Best Local Similarity   27.5%;  Pred. No. 0.43;
  Matches   56;  Conservative   31;  Mismatches   64;  Indels   53;  Gaps   12;

Qy        129 PPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGY----PGCPYPPA 184
              ||  |||:|    ||     ||| | |  |:  |: ||| || |       |  | ||
Db        235 PPPPSRPSVEVP-PP-----PPNRMYPPPPPALPSSAPSGPPPPPPSVLGVGPVAPPPP 288

Qy        185 GPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGAQAEL 244
              | |       |  ||     || ||     |  |   :|: |::|       :|||
Db        289 PPPP-------PPGPPPPPGLPS-DGDHQVPT-TAGNKAALLDQIR-------EGAQ--- 329

Qy        245 NALKRTEEDLK----KGHQKLEEMVTRLDQ--EVAE--------------VDKNIELLK 283
                ||: |:: :     |   | : : :  |   ||:                |   :|:::
Db        330 --LKKVEQNSRPVSCSGRDALLDQIRQGIQLKSVADGQESTPPTPAPTSGIVGALMEVMQ 387

Qy        284 KKDEELSSALEKMENQSENNDIDE 307
              |: : : |: :: |:: :   | ::
Db        388 KRSKAIHSS-DEDEDEDDEEDFED 410
```

```
RESULT 9
US-09-284-320-2
; Sequence 2, Application US/09284320
; Publication No. US20030092175A1
; GENERAL INFORMATION:
;  APPLICANT:  Kato, Seishi et al.
;  TITLE OF INVENTION:  HUMAN PROTEINS HAVING TRANSMEMBRANE DOMAINS AND DNAs
;  TITLE OF INVENTION:  ENCODING THESE PROTEINS
;  FILE REFERENCE:  GIN-6705CPUS
;  CURRENT APPLICATION NUMBER: US/09/284,320
;  CURRENT FILING DATE:  1999-06-21
;  PRIOR APPLICATION NUMBER:  JP 8-301429
;  PRIOR FILING DATE:  1996-11-13
;  PRIOR APPLICATION NUMBER:  PCT/JP97/04056
;  PRIOR FILING DATE:  1997-11-07
;  NUMBER OF SEQ ID NOS:  91
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 371
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-284-320-2

  Query Match             6.2%;  Score 124.5;  DB 9;  Length 371;
  Best Local Similarity   35.8%;  Pred. No. 0.51;
  Matches   34;  Conservative    2;  Mismatches   28;  Indels   31;  Gaps    6;

Qy        126 GEEPPVFSRPTVSASYPPYTATGPPNTSY----------MPGMPSGISAYP-SGYPPNP 173
              | :||:          |||     |   |             || | | | ||  |||  |
Db         24 GPQPPM----------PPYAQPPYPGAPYPQPPFQPSPYGQPGYPHGPSPYPQGGYPQGP 73

Qy        174 ---SGYPGCPYP----PAGPYPATTSSQ--YPSQP 199
                 |||  |||     | ||||     |  ||  |
Db         74 YPQGGYPQGPYPQEGYPQGPYPQGGYPQGPYPQSP 108


RESULT 10
US-10-197-666A-4
; Sequence 4, Application US/10197666A
; Publication No. US20030092037A1
; GENERAL INFORMATION:
;  APPLICANT: ASAHI KASEI KABUSIKI KAISYA
;  TITLE OF INVENTION: Elk1 phosphorylation related gene
;  FILE REFERENCE: PH-1548US
;  CURRENT APPLICATION NUMBER: US/10/197,666A
;  CURRENT FILING DATE:  2002-11-18
;  PRIOR APPLICATION NUMBER: JP 2001-218204
;  PRIOR FILING DATE: 2001-07-18
;  PRIOR APPLICATION NUMBER: JP 2001-263450
;  PRIOR FILING DATE: 2001-08-31
;  PRIOR APPLICATION NUMBER: JP 2002-012176
;  PRIOR FILING DATE: 2002-01-21
;  PRIOR APPLICATION NUMBER: US 60/305,884
;  PRIOR FILING DATE: 2001-07-18
;  PRIOR APPLICATION NUMBER: US 60/316,304
;  PRIOR FILING DATE: 2001-09-04
;  PRIOR APPLICATION NUMBER: US 60/350,027
;  PRIOR FILING DATE: 2002-01-23
;  NUMBER OF SEQ ID NOS: 156
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 371
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-197-666A-4

  Query Match             6.2%;  Score 124.5;  DB 9;  Length 371;
  Best Local Similarity   35.8%;  Pred. No. 0.51;
  Matches   34;  Conservative    2;  Mismatches   28;  Indels   31;  Gaps    6;

Qy        126 GEEPPVFSRPTVSASYPPYTATGPPNTSY----------MPGMPSGISAYP-SGYPPNP 173
              | :||:          |||     |   |             || | | | ||  |||  |
Db         24 GPQPPM----------PPYAQPPYPGAPYPQPPFQPSPYGQPGYPHGPSPYPQGGYPQGP 73

Qy        174 ---SGYPGCPYP----PAGPYPATTSSQ--YPSQP 199
                 |||  |||     | ||||     |  ||  |
Db         74 YPQGGYPQGPYPQEGYPQGPYPQGGYPQGPYPQSP 108


RESULT 11
US-09-771-161A-192
; Sequence 192, Application US/09771161A
; Patent No. US20020110811A1
; GENERAL INFORMATION:
;  APPLICANT: LEVINE, et al.
;  TITLE OF INVENTION: VARIANTS Of PROTEIN KINASES
;  FILE REFERENCE: 802620-2005.1
;  CURRENT APPLICATION NUMBER: US/09/771,161A
;  CURRENT FILING DATE:  2001-01-26
;  PRIOR APPLICATION NUMBER: 09/724,676
;  PRIOR FILING DATE: 2000-11-28
;  PRIOR APPLICATION NUMBER: 136776
;  PRIOR FILING DATE: 2000-06-15
;  PRIOR APPLICATION NUMBER: 135619
;  PRIOR FILING DATE: 2000-04-12
;  NUMBER OF SEQ ID NOS: 273
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 192
;   LENGTH: 887
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-771-161A-192

  Query Match             6.1%;  Score 121;  DB 10;  Length 887;
  Best Local Similarity   23.8%;  Pred. No. 2.8;
  Matches   56;  Conservative   41;  Mismatches   86;  Indels   52;  Gaps   10;

Qy        186 PYPATTSSQYPSQPPVTTVGPSRDGTISEDTI------RASLISAVSDKLRWRM---KEE 236
              | |: :     | || :      ||    |  :      |||  | ::: | | :    |:
Db        449 PLPSVS-----SPPPASKTKEVPDGENLEQDLCTFLISRASKNSTLANYLYWYVIVECED 503

Qy        237 MDGAQAE-------LNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELL------- 282
              |  | :        || ::|  : | || : :  | : |  :   ||: : |:
Db        504 QDTQQRDPKTHEMYLNVMRRFSQALLKGDKSVRVMRSLLAAQQTFVDRLVHLMKAVQRES 563

Qy        283 ---KKKDEELSSALEKMENQSEN-NDIDEVIIP-------------TAPLYKQIL---NL 322
                 |||:| | : |   :|:  | :|:: : :|             || |:|  |    |
Db        564 GNRKKKNERLQALLG--DNEKMNLSDVELIPLPLEPQVKIRGIIPETATLFKSALMPAQL 621

Qy        323 Y--AEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRKQFQLRALMQKARKTA 375
              :   |:      ||  |: ||:  : | :     :|| ::   |:    |   |:
Db        622 FFKTEDGGKYPVIFKHGDDLRQDQLILQIISLMDKLLRKENLDLKLTPYKVLATS 676


RESULT 12
US-09-771-161A-193
; Sequence 193, Application US/09771161A
; Patent No. US20020110811A1
; GENERAL INFORMATION:
;  APPLICANT: LEVINE, et al.
;  TITLE OF INVENTION: VARIANTS Of PROTEIN KINASES
;  FILE REFERENCE: 802620-2005.1
;  CURRENT APPLICATION NUMBER: US/09/771,161A
;  CURRENT FILING DATE:  2001-01-26
;  PRIOR APPLICATION NUMBER: 09/724,676
;  PRIOR FILING DATE: 2000-11-28
;  PRIOR APPLICATION NUMBER: 136776
;  PRIOR FILING DATE: 2000-06-15
;  PRIOR APPLICATION NUMBER: 135619
;  PRIOR FILING DATE: 2000-04-12
;  NUMBER OF SEQ ID NOS: 273
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 193
;   LENGTH: 887
;   TYPE: PRT
```

```
;   ORGANISM: Homo sapiens
US-09-771-161A-193

  Query Match             6.1%;  Score 121;  DB 10;  Length 887;
  Best Local Similarity   23.8%;  Pred. No. 2.8;
  Matches   56;  Conservative   41;  Mismatches   86;  Indels   52;  Gaps   10;

Qy        186 PYPATTSSQYPSQPPVTTVGPSRDGTISEDTI------RASLISAVSDKLRWRM---KEE 236
              | |: :     | || :      ||   |  :      |||  | ::: | | :    |:
Db        449 PLPSVS-----SPPPASKTKEVPDGENLEQDLCTFLISRASKNSTLANYLYWYVIVECED 503

Qy        237 MDGAQAE-------LNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELL------- 282
               |  | :       || ::|  : | || : :  | : |  :   ||: : |:
Db        504 QDTQQRDPKTHEMYLNVMRRFSQALLKGDKSVRVMRSLLAAQQTFVDRLVHLMKAVQRES 563

Qy        283 ---KKKDEELSSALEKMENQSEN-NDIDEVIIP-------------TAPLYKQIL---NL 322
                 |||:| | : |   :|:  | :|:: : :|             || |:|  |    |
Db        564 GNRKKKNERLQALLG--DNEKMNLSDVELIPLPLEPQVKIRGIIPETATLFKSALMPAQL 621

Qy        323 Y--AEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRKQFQLRALMQKARKTA 375
              :   |:          ||  |: ||:  : | :     :|| ::   |:    |   |:
Db        622 FFKTEDGGKYPVIFKHGDDLRQDQLILQIISLMDKLLRKENLDLKLTPYKVLATS 676


RESULT 13
US-09-823-240-10
; Sequence 10, Application US/09823240
; Patent No. US20020048813A1
; GENERAL INFORMATION:
;  APPLICANT: Frank B. Gertler
;  APPLICANT:  James E.Bear
;  APPLICANT:  Jurgen Wehland
;  APPLICANT:  Joseph Loureio
;  TITLE OF INVENTION: Methods and Products for Regulating Cell
;  TITLE OF INVENTION:  Motility
;  FILE REFERENCE: M0656/7064 (HCL)
;  CURRENT APPLICATION NUMBER: US/09/823,240
;  CURRENT FILING DATE:  2001-03-30
;  PRIOR APPLICATION NUMBER: 60/194,564
;  PRIOR FILING DATE: 2000-04-03
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 10
;   LENGTH: 380
;   TYPE: PRT
;   ORGANISM: Homo sapien
US-09-823-240-10

  Query Match             6.0%;  Score 119;  DB 10;  Length 380;
  Best Local Similarity   21.2%;  Pred. No. 1.3;
  Matches   72;  Conservative   42;  Mismatches  124;  Indels  102;  Gaps   13;

Qy         33 VFNDGSSRELVNLTGTIPVRY-RGNIYNIPICLWLLDTYPYNPPICFVKPTSSMTIKTGK 91
              :::||: | |    || |  : |  ||:                          ||::     |:
Db         15 LYDDGNKRWLP--AGTGPQAFSRVQIYH------------------NPTANSFRVVGR 52

Qy         92 HVDANGKIYL-------------PYLHDWKHPR--------------------SELLEL 117
               :  : :: :             |  | |:  |                     :  ||
Db         53 KMQPDQQVVINCAIVRGVKYNQATPNFHQWRDARQVWGLNFGSKEDAAQFAAGMASALEA 112

Qy        118 IQIMIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPS-----------GISAYP 166
              ::       |  ||  : || |     |             || ||           |  | |
Db        113 LE-----GGGPPPPPALPTWSVPNGPSPEEVEQQKRQQPG-PSEHIERRVSNAGGPPAPP 166

Qy        167 SGYPPNPSGYPGCPYPPAGPYPATTSSQYPSQ-------PPVTTVGPSRDGTISEDTIRA 219
              :| || | | |  | ||  | |    |  |:        ||     |: |
Db        167 AGGPPPPPGPPPPPGPP--PPPGLPPSGVPAAAHGAGGGPPPAPPLPAAQGPGGGGAGAP 224

Qy        220 SLISAVSD-KLRWRMKEEM----------------DGAQAELNA-LKRTEEDLKKGHQK 260
               | :|::  |||   |:|                    |   |:|| | |  :  : | :
Db        225 GLAAAIAGAKLRKVSKQEEASGGPTAPKAESGRSGGGGLMEEMNAMLARRRKATQVGEKT 284

Qy        261 LEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 300
```

```
              ::     ::  | |    | :::  |: |: | :|:: |
Db        285 PKDESANQEEPEARVPAQSESVRRPWEKNSTTLPRMKSSS 324


RESULT 14
US-10-072-036-125
; Sequence 125, Application US/10072036
; Publication No. US20030082564A1
; GENERAL INFORMATION:
;  APPLICANT: Ole THASTRUP
;  APPLICANT:  Sara BJRON
;  APPLICANT:  Soren TULLIN
;  APPLICANT:  Kasper ALMHOLT
;  APPLICANT:  Kurt SCUDDER
;  TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To A
;  TITLE OF INVENTION:  On A Cellular Response
;  FILE REFERENCE: 3759-0120P
;  CURRENT APPLICATION NUMBER: US/10/072,036
;  CURRENT FILING DATE:  2002-09-13
;  PRIOR APPLICATION NUMBER: 09/417,197
;  PRIOR FILING DATE: 1999-10-07
;  NUMBER OF SEQ ID NOS: 143
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 125
;   LENGTH: 635
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: EGFP-VASP fusion
US-10-072-036-125

  Query Match             6.0%;  Score 119;  DB 9;  Length 635;
  Best Local Similarity   21.2%;  Pred. No. 2.5;
  Matches   72;  Conservative   42;  Mismatches  124;  Indels  102;  Gaps   13;

Qy         33 VFNDGSSRELVNLTGTIPVRY-RGNIYNIPICLWLLDTYPYNPPICFVKPTSSMTIKTGK 91
              :::||: | |    || |  : |  ||:                          ||::     |:
Db        270 LYDDGNKRWLP--AGTGPQAFSRVQIYH------------------NPTANSFRVVGR 307

Qy         92 HVDANGKIYL-------------PYLHDWKHPR--------------------SELLEL 117
               :  : :: :             |  | |:  |                     :  ||
Db        308 KMQPDQQVVINCAIVRGVKYNQATPNFHQWRDARQVWGLNFGSKEDAAQFAAGMASALEA 367

Qy        118 IQIMIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPS-----------GISAYP 166
              ::       |  ||  : || |     |             || ||           |  | |
Db        368 LE-----GGGPPPPPALPTWSVPNGPSPEEVEQQKRQQPG-PSEHIERRVSNAGGPPAPP 421

Qy        167 SGYPPNPSGYPGCPYPPAGPYPATTSSQYPSQ-------PPVTTVGPSRDGTISEDTIRA 219
              :| || | | |  | ||  | |    |  |:        ||     |: |
Db        422 AGGPPPPPGPPPPPGPP--PPPGLPPSGVPAAAHGAGGGPPPAPPLPAAQGPGGGGAGAP 479

Qy        220 SLISAVSD-KLRWRMKEEM----------------DGAQAELNA-LKRTEEDLKKGHQK 260
               | :|::  |||   |:|                    |   |:|| | |  :  : | :
Db        480 GLAAAIAGAKLRKVSKQEEASGGPTAPKAESGRSGGGGLMEEMNAMLARRRKATQVGEKT 539

Qy        261 LEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 300
              ::     ::  | |    | :::  |: |: | :|:: |
Db        540 PKDESANQEEPEARVPAQSESVRRPWEKNSTTLPRMKSSS 579


RESULT 15
US-10-083-357-1277
; Sequence 1277, Application US/10083357
; Publication No. US20030054370A1
; GENERAL INFORMATION:
;  APPLICANT: Qiandong Zeng et al.
;  TITLE OF INVENTION: Systemic Discovery of New Genes
;  FILE REFERENCE: 032796-090
;  CURRENT APPLICATION NUMBER: US/10/083,357
;  CURRENT FILING DATE:  2002-02-27
;  NUMBER OF SEQ ID NOS: 1346
; SEQ ID NO 1277
;   LENGTH: 92
```

; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-083-357-1277

Query Match 5.9%; Score 118.5; DB 9; Length 92;
Best Local Similarity 33.3%; Pred. No. 0.22;
Matches 32; Conservative 20; Mismatches 37; Indels 7; Gaps 2;

```
Qy        283 KKKDEELSSALEKMENQSENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRR 342
              : | :: ||      :  :: |:: : :       |: || |:: |: |||  |   | |
Db          4 RAKVQQFSST-----SHVDDEDVNSIAVAKTDGLNQLYNLVAQDYALTDTIECLSRMLHR 58

Qy        343 GVIDLDVFLKHVRLLSRKQFQLRALMQKARKTAGLS 378
              | | || |:|  | |:|:|| :|  :|  | |: ||
Db         59 GTIPLDTFVKQGRELARQQFLVRWHIQ--RITSPLS 92
```

Search completed: June 3, 2003, 14:22:34
Job time : 31.5401 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: June 3, 2003, 14:06:00 ; Search time 26.0342 Seconds
(without alignments)
1406.890 Million cell updates/sec

Title: US-09-804-690-2
Perfect score: 1996
Sequence: 1 MMSKYKYRDLTVRQTVNVIA.....FQLRALMQKARKTAGLSDLY 381

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	627.5	31.4	404	2	H88650	protein C09G12.9 [
2	396	19.8	83	2	I48283	gene CC2 protein -
3	151	7.6	169	2	T34520	hypothetical prote
4	142	7.1	212	2	S74288	hypothetical prote
5	141.5	7.1	485	2	T37550	hypothetical coile
6	137.5	6.9	503	1	LURB11	annexin XI - rabbi
7	137	6.9	488	1	LUHU7	annexin VII, long
8	135.5	6.8	463	2	S29170	annexin VII - mous
9	135.5	6.8	827	2	T39608	zinc finger transc
10	133	6.7	198	2	D70509	hypothetical prote
11	133	6.7	505	2	S72273	actin-depolymerizi
12	130	6.5	1006	2	T42731	atrophin-1 related
13	129	6.5	669	2	T28754	hypothetical prote
14	128	6.4	505	2	A53152	annexin XI - human
15	127.5	6.4	1281	2	JC5368	dynactin 1 - mouse
16	127	6.4	279	2	T05421	hypothetical prote
17	124	6.2	553	2	G83385	hypothetical prote
18	123.5	6.2	1181	2	C86349	F8K7.4 protein - A
19	122.5	6.1	139	2	H84809	hypothetical prote
20	122.5	6.1	503	1	LUBO11	annexin XI form A
21	122.5	6.1	505	1	S23447	annexin XI form B
22	122.5	6.1	1902	2	C97702	cell surface antig
23	122	6.1	792	2	T49989	hypothetical prote
24	122	6.1	978	2	A70387	conserved hypothet
25	121.5	6.1	214	2	T10737	extensin-like cell
26	121.5	6.1	214	2	T09854	proline-rich cell
27	121	6.1	887	1	S57219	1-phosphatidylinos
28	120.5	6.0	338	2	I53043	transforming prote
29	120.5	6.0	729	2	E70803	hypothetical prote
30	120	6.0	678	2	H88187	protein C18H9.8 [i
31	120	6.0	1605	2	T31435	DNA-directed RNA p
32	119.5	6.0	815	2	B56708	extracellular sign
33	119.5	6.0	1189	2	T42726	guanine nucleotide
34	119.5	6.0	1613	2	S39059	protein BRG1 - hum
35	119.5	6.0	1733	1	RNBY2L	DNA-directed RNA p
36	119.5	6.0	1802	2	H88444	protein C26E6.12 [
37	119	6.0	380	2	S51797	vasodilator-stimul
38	118.5	5.9	384	2	S51796	vasodilator-stimul
39	118.5	5.9	1647	2	S45252	SNF2beta protein -
40	118	5.9	929	2	C96623	hypothetical prote
41	117.5	5.9	262	2	A54889	IgE-binding protei
42	117.5	5.9	1184	2	G01763	atrophin-1 - human
43	117	5.9	437	2	T14192	extensin homolog T
44	117	5.9	609	2	S46019	YSW1 protein - yea
45	117	5.9	964	2	T21865	hypothetical prote

ALIGNMENTS

RESULT 1
H88650
protein C09G12.9 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C;Accession: H88650
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating bio
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www_sanger.ac.uk/Projects/C_
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;
A;Accession: H88650
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-404 <STO>
A;Cross-references: GB:chr_IV; PIDN:AAC25822.1; PID:g3294495; GSPDB:GN00022; CESP:C09
C;Genetics:
A;Gene: C09G12.9
A;Map position: 4

Query Match 31.4%; Score 627.5; DB 2; Length 404;
Best Local Similarity 36.1%; Pred. No. 2.2e-30;
Matches 145; Conservative 78; Mismatches 134; Indels 45; Gaps 10;

```
Qy          6 KYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNIPICLW 65
              || |   :  :  :: :||| |  |:::| ||  |    | ||||| |:|  ||||: ::
Db         16 KYADSAKKDIIGALSQFKDLSPGTDTFMFPDGKRRTAFRLKGTIPVYYKGACYNIPVTVY 75

Qy         66 LLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQIMIVIF 125
             | ||:||  |||:| |||:|     :||:  ||::|||::|:  |  :|  |:|::
Db         76 LWDTHPYYAPICYVNPTSTM---ESEHVNKEGKVFLPYLNEWRFPGYDLSGLLQMIFARS 132

Qy        126 GEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSG-----YPPNPSGYPGC- 179
                 :  |  ::    ::|  |  |  | ||   : ||:|     |||: : ||
Db        133 AANSATNASATNPSAGSSASSTPTPYPSSQPTMP---TPYPTGSGAAPYPPSSTPYPSAG 189

Qy        180 -----PY---PPAGPYP--ATTSSQYPS----------QPPV---TTVGPSRDGTISED 215
                 ||   | : |||  |: :| |||            |||   |:|  |  |||  |
Db        190 AMGYNPYMNVPQSTPYPMGASGASPYPSASSNPAPPPPRPPPVTAQTSVSSSSGGTIQAD 249

Qy        216 TIRASLISAVSDKLRWRMKEEMDGAQAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEV 275
              |||||::||| :|:| :::| |    ||: ::: | ::|::| |||: |:  |: : : :
Db        250 TIRASVMSAVEEKIRAKLRERMGTNSAEMASIRTTSDELREGQQKLKRMLEELETQRSSL 309

Qy        276 DKNIELLKKKDEELSSALEKMENQSENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFY 335
              |:   |  ||: ||      ::   ||| |    ||::||:  ||::
Db        310 QTACEIYTAKKAELAKALSD-AGGTDAPPIDEAIDAAFPLHRQIVLNYAKDLTC------ 362

Qy        336 LGEALRRGVIDLDVFLKHVRLLSRKQFQLRALMQKARKTAGL 377
              ::|::  | |  :|:||| :||:||  || ||| |:||||
Db        363 --QSLKKRQITLAEYLRHVRDVSREQFIYRATMQKCRRTAGL 402
```

RESULT 2
I48283
gene CC2 protein - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 05-Nov-1999
C;Accession: I48283
R;Maucuer, A.; Camonis, J.H.; Sobel, A.
Proc. Natl. Acad. Sci. U.S.A. 92, 3100-3104, 1995
A;Title: Stathmin interaction with a putative kinase and coiled-coil-forming protein dom
A;Reference number: I48282; MUID:95241452; PMID:7724523
A;Accession: I48283
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-83 <RES>
A;Cross-references: EMBL:X82319; NID:g791075; PIDN:CAA57762.1; PID:g791076
C;Genetics:
A;Gene: CC2

Query Match 19.8%; Score 396; DB 2; Length 83;
Best Local Similarity 98.8%; Pred. No. 2.1e-17;
Matches 81; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      232 RMKEEMDGAQAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSS 291
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 RMKEEMDGAQAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVGKNIELLKKKDEELSS 60

Qy      292 ALEKMENQSENNDIDEVIIPTA 313
            |||||||||||||||||||||||
Db       61 ALEKMENQSENNDIDEVIIPTA 82
```

RESULT 3
T34520
hypothetical protein DKFZp564J157.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T34520
R;Bloecker, H.; Boecher, M.; Brandt, P.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, October 1999
A;Reference number: Z21539
A;Accession: T34520
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-169 <BLO>
A;Cross-references: EMBL:AL122042
A;Experimental source: fetal brain; clone DKFZp564J157
C;Genetics:
A;Note: DKFZp564J157.1

Query Match 7.6%; Score 151; DB 2; Length 169;
Best Local Similarity 28.7%; Pred. No. 0.023;
Matches 39; Conservative 15; Mismatches 32; Indels 50; Gaps 6;

```
Qy      135 PTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNP---SGYPGC----PYPPAGPY 187
            | ::  :||     |      || | |  |:||| ||:|    |||||    ||||  |
Db       51 PPINPPFPPGPCPPP------PGAPHGNPAFPSGGPPHPVPQPGYPGCQPLGPYPPPYPP 104

Qy      188 PATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGAQAELNAL 247
            ||      |  |||  : |   |           : : |
Db      105 PA------PGIPPVNPLAPGMVGP----------AVIVD-------------------- 127

Qy      248 KRTEEDLKKGHQKLEE 263
            |: :: :|| |:|: :
Db      128 KKMQKKMKKAHKKMHK 143
```

RESULT 4
S74288
hypothetical protein YCL008c - yeast (Saccharomyces cerevisiae)
C;Species: Saccharomyces cerevisiae
C;Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 19-Apr-2002
C;Accession: S74288; S19410
R;Wedler, H.; Wambutt, R.
submitted to the Protein Sequence Database, September 1996
A;Reference number: S74288
A;Accession: S74288
A;Molecule type: DNA
A;Residues: 1-212 <WED>
A;Cross-references: EMBL:X59720; NID:g1907116; PIDN:CAA42351.1; PID:e309039; PID:g190
R;Oliver, S.G.; Anwar, R.; Brown, A.; Gent, M.E.; Indge, K.J.; James, C.M.; Stateva,
submitted to the Protein Sequence Database, March 1992
A;Reference number: S19337
A;Accession: S19410
A;Molecule type: DNA
A;Residues: 1-90,'GEYS',95,'TA',98-99,'AQVSTP',106,'TAIASTSTTSTGI' <OLI>
A;Cross-references: EMBL:X59720; MIPS:YCL008c
A;Note: this sequence has been revised in reference S74288
C;Genetics:
A;Gene: SGD:STP22
A;Cross-references: SGD:S0000514
A;Map position: 3L
A;Note: YCL008c

Query Match 7.1%; Score 142; DB 2; Length 212;
Best Local Similarity 23.7%; Pred. No. 0.1;
Matches 57; Conservative 37; Mismatches 99; Indels 48; Gaps 11;

```
Qy       64 LWLLDTYPYNPPIC------FVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLEL 117
            :|:   ||  ||         |   | | ::   :::|:|| | || || |      |: :
Db        1 MWVPSMYPVKPPFISINLENFDMNTISSSLPIQEYIDSNGWIALPILHCWDPAAMNLIMV 60

Qy      118 IQIMIVIFGEEPPVFSRPTVSASYPPYTAT---GPPNTSYMPGMPSGISAYPSGYPPNPS 174
            :| :: :   |||    |    | ||   |      ||  :|  |      |  || |
Db       61 VQELMSLL-HEPPQDQAP----SLPPKPNTQLQQEQNTPPLPPKPKS----PHLKPPLP- 110

Qy      175 GYPGCPYPPAGPYPATTSSQYPSQPPVTTVGPSRD-------GTISEDTIRASLISAVSD 227
                ||  | ||: :         | : |:          |:  :  |   :  |:|
Db      111 -------PPPPPQPASNALDLMDMDN-TDISPTNHHEMLQNLQTVVNELYRED-VDYVAD 161

Qy      228 KLRWR---MKEEMDGAQAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKK 284
            |:  |   |:|          :: |  | :      |  :   ::| :  ::  |::| :
Db      162 KILTRQTVMQE----------SIARFHEIIAIDKNHLRAVEQAIEQTMHSLNAQIDVLNR 211

Qy      285 K 285
            |
Db      212 K 212
```

RESULT 5
T37550
hypothetical coiled-coil protein [imported] - fission yeast (Schizosaccharomyces pomb
C;Species: Schizosaccharomyces pombe
C;Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 08-Sep-2000
C;Accession: T37550
R;Connor, R.; Churcher, C.M.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, October 1999
A;Reference number: Z21723
A;Accession: T37550
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-485 <CON>
A;Cross-references: EMBL:AL132714; PIDN:CAB59797.1; GSPDB:GN00066; SPDB:SPAC11H11.01
A;Experimental source: strain 972h-; cosmid c11H11
C;Genetics:
A;Gene: SPDB:SPAC11H11.01
A;Map position: 1

Query Match 7.1%; Score 141.5; DB 2; Length 485;
Best Local Similarity 21.2%; Pred. No. 0.31;
Matches 90; Conservative 54; Mismatches 135; Indels 145; Gaps 17;

```
Qy       63 CLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRS--ELLELIQI 120
            |: :   ||  |||         :  | |||      |     |  ||: ||   :  |:|
Db       79 CITVRSYYPSQPPI-----VQLLDEKGGKH-----KYTSLLLQLWKNERSVFNIYRLVQA 128
```

```
Qy        121 MIVIFGEEPPVFSRPTVSASYPPYTATGPPNT-------------SYMPGMPSG------ 161
              :|         | |   |   ||   |   ||                 :|  |
Db        129 LI------KQDFEREHTS---PPELPTKLVNTIEKLKVKEENEAPPVIPAKPFSSSSEQH 179

Qy        162 ---ISAYPSGYPPNP-----------------SGYPGCPYP-PAGPYPATT--------- 191
                 : | ||  || |                 | :    :     |: ||:
Db        180 FRKVPALPSKLPPKPLKITANSSLGQETNSNSSSFQSTLFSLNTAPFSATSQQLVHDSVS 239

Qy        192 ----SSQYPSQPPVTTVGPSRDGTISEDT------------------IRASLISAVSDK 228
                  ||  |:| |:       :  |::||                  :  ::  ||
Db        240 LRRPSSNIPAQKPIPPKPEQNEIIITKDTPSLKDKYSKPALLPQKPKVSKGQIVQQVSVF 299

Qy        229 LRWRMKEEMDGAQAELNALKRTEEDLKKGHQKL-------------------------- 261
                 :  |    :|: || :    |   ||  :|
Db        300 STGKKIE----SQSLLNLIDTDIETPLKGSSELLYSEDFKPNVDPVKIQQILHKQNKIIE 355

Qy        262 EEMVTR-------------LDQE---VAEVDKNIE----LLKKKDEELSSALEKMENQSE 301
              |: :::              ||||   :  : ||||    :| |:  :   ||:|::|  :
Db        356 EKWISQIRISKNLEVKQRLLDQERHALETLAKNIENNRFILGKRRRKAREALQKLDNLKD 415

Qy        302 NNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRKQ 361
               :  :  |||:    |    |  ::  ::: |  | :||    |    :|| ::||:|:|
Db        416 LSVQELFIIPSERELK-YYELKRKDEKLDEGIRALNQALHHESIMPASWLKGIKLLARQQ 474

Qy        362 FQLR 365
              | :|
Db        475 FLIR 478
```

RESULT 6
LURB11
annexin XI - rabbit
N;Alternate names: calcyclin-associated annexin protein CAP-50
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Jun-2000
C;Accession: JH0694; PH0950; A38250; PS0263
R;Tokumitsu, H.; Mizutani, A.; Muramatsu, M.; Yokota, T.; Arai, K.; Hidaka, H.
Biochem. Biophys. Res. Commun. 186, 1227-1235, 1992
A;Title: Molecular cloning of rabbit CAP-50, a calcyclin-associated annexin protein.
A;Reference number: JH0694; MUID:92378579; PMID:1380798
A;Accession: JH0694
A;Molecule type: mRNA
A;Residues: 1-503 <TOK>
A;Cross-references: DDBJ:D10883; NID:g471147; PIDN:BAA01705.1; PID:g471148
A;Experimental source: lung
A;Accession: PH0950
A;Molecule type: protein
A;Residues: 104-141;213-231;254-262;270-280;285-309;319-337;429-448;478-492 <TK2>
R;Tokumitsu, H.; Mizutani, A.; Minami, H.; Kobayashi, R.; Hidaka, H.
J. Biol. Chem. 267, 8919-8924, 1992
A;Title: A calcyclin-associated protein is a newly identified member of the Ca(2+)/phosp
A;Reference number: A38250; MUID:92250478; PMID:1533622
A;Accession: A38250
A;Molecule type: protein
A;Residues: 104-141;213-223,'X',225-231;254-263;271-280;285-291,'X',293-300,'X',302-309;
C;Comment: This protein binds specifically to calcyclin in a Ca2+ dependent manner.
C;Superfamily: annexin VII; annexin repeat homology
C;Keywords: calcium binding; duplication; endonexin fold; glycoprotein; phospholipid bin
F;201-272/Domain: annexin repeat homology <AX1>
F;212-228/Region: endonexin fold #status predicted
F;273-344/Domain: annexin repeat homology <AX2>
F;284-300/Region: endonexin fold #status predicted
F;356-428/Domain: annexin repeat homology <AX3>
F;368-384/Region: endonexin fold #status predicted
F;432-503/Domain: annexin repeat homology <AX4>
F;443-459/Region: endonexin fold #status predicted
F;58/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match   6.9%; Score 137.5; DB 1; Length 503;
  Best Local Similarity   22.7%; Pred. No. 0.57;
  Matches   75; Conservative   42; Mismatches 117; Indels   97; Gaps   15;

```
Qy        125 FGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSG-YPGCPYP- 182
              ||: ||  ::|:|    |||    ||  :   |:||        ||| |    || | |
Db         89 FGQPPP--TQPSV----PPYGVYPPPGGNPPSGVPS--------YPPFPGAPVPGQPMPP 134

Qy        183 ----PAGPYPATTSSQYPSQPPVTTVG----PSRDGTISEDTI----------RASLIS 223
                  | ||||       || | ||   |    ||  |    |:                | ::
Db        135 PGHQPPGPYPGQLPVTYPGQSPVPPPGQQPMPSYPGYPGSGTVTPAVPPVQFGNRGTITD 194

Qy        224 AV-------SDKLRWRMKEEMDGAQAELNAL-----KRTEEDLKK-----GHQKLEEMVT 266
              |        :: ||  ||       || :: |     |: :: |       |   :::: :
Db        195 ASGFDPLRDAEVLRKAMKGFGTDEQAIIDCLGSRSNKQRQQILLSFKTAYGKDLIKDLKS 254

Qy        267 RLD------------------QEVAEVDKN--------IELLKKKDEELSSALEKMENQ 299
              |                    |: |  |          ||:|  :  |    | |
Db        255 ELSGNFEKTILALMKTPILFDAYEIKEAIKGAGTDEACLIEILASRSNEHIRELNKAYKT 314

Qy        300 SENNDIDEVI-IPTAPLYKQILNLYAEENAIEDT-------------IFYLGEALRRGVI 345
                  ::| |   |:  ::::|   :: |  | |             ::  ||      :
Db        315 EFKKTLEEAIRSDTSGHFQRLLISLSQGNRDESTNVDMSLVQRDVQELYAAGE----NRL 370

Qy        346 DLDVFLKHVRLLSRKQFQLRALMQKARKTAG 376
               |    :  | || :  | |:  : ::  |
Db        371 GTDESKFNAVLCSRSRAHLVAVFNEYQRMTG 401
```

RESULT 7
LUHU7
annexin VII, long form - human
N;Alternate names: synexin
N;Contains: annexin VII, long form; annexin VII, short form
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1992 #sequence_revision 26-Jan-1996 #text_change 22-Jun-1999
C;Accession: A54467; A32554; A39513; B39513
R;Shirvan, A.; Srivastava, M.; Wang, M.G.; Cultraro, C.; Magendzo, K.; McBride, O.W.;
Biochemistry 33, 6888-6901, 1994
A;Title: Divergent structure of the human synexin (annexin VII) gene and assignment t
A;Reference number: A54467; MUID:94264005; PMID:7515686
A;Accession: A54467
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-488 <SHI>
R;Burns, A.L.; Magendzo, K.; Shirvan, A.; Srivastava, M.; Rojas, E.; Alijani, M.R.; P
Proc. Natl. Acad. Sci. U.S.A. 86, 3798-3802, 1989
A;Title: Calcium channel activity of purified human synexin and structure of the huma
A;Reference number: A32554; MUID:89264510; PMID:2542947
A;Accession: A32554
A;Molecule type: mRNA
A;Residues: 1-145,168-488 <BUR>
A;Cross-references: EMBL:J04543; NID:g338243; PIDN:AAA36616.1; PID:g338244
R;Magendzo, K.; Shirvan, A.; Cultraro, C.; Srivastava, M.; Pollard, H.B.; Burns, A.L.
J. Biol. Chem. 266, 3228-3232, 1991
A;Title: Alternative splicing of human synexin mRNA in brain, cardiac, and skeletal m
A;Reference number: A39513; MUID:91131630; PMID:1825209
A;Accession: A39513
A;Molecule type: mRNA
A;Residues: 137-145,168-176 <MAG>
A;Cross-references: EMBL:J05732
A;Accession: B39513
A;Molecule type: mRNA
A;Residues: 137-176 <MA2>
C;Comment: Annexins undergo reversible, calcium-dependent binding to membrane phospho
derstood.
C;Comment: The long form of annexin VII is more prevalent in brain, heart, and skelet
C;Genetics:
A;Gene: GDB:ANX7
A;Cross-references: GDB:369042; OMIM:186360
A;Map position: 10q21.1-10q21.2
A;Introns: 17/3; 87/1; 124/1; 145/3; 167/3; 202/2; 233/3; 271/3; 328/3; 385/3; 411/1;
C;Superfamily: annexin VII; annexin repeat homology
C;Keywords: alternative splicing; calcium binding; calcium channel; duplication; endo
F;1-488/Product: annexin VII, long form #status predicted <MAT>
F;1-145,168-488/Product: annexin VII, short form #status predicted <MA3>
F;188-259/Domain: annexin repeat homology <AX1>

F;199-215/Region: endonexin fold #status predicted
F;260-331/Domain: annexin repeat homology <AX2>
F;271-287/Region: endonexin fold #status predicted
F;343-415/Domain: annexin repeat homology <AX3>
F;355-371/Region: endonexin fold #status predicted
F;419-488/Domain: annexin repeat homology <AX4>
F;430-446/Region: endonexin fold #status predicted

Query Match 6.9%; Score 137; DB 1; Length 488;
Best Local Similarity 20.7%; Pred. No. 0.59;
Matches 85; Conservative 34; Mismatches 92; Indels 200; Gaps 17;

```
Qy      140 SYPPYTATG-PPNTSYMPG------MPSGISAYPSGYP---------------------- 170
            ||| |  || ||   | |        |||   ||||:|
Db        2 SYPGYPPTGYPPFPGYPPAGQESSFPPSGQYPYPSGFPPMGGGAYPQVPSSGYPGAGGYP 61

Qy      171 -----PNPSGYPGCPYPPAGP-YPATTSSQ-------------YPSQPPVTTVG------ 205
                | | |||| | |   | ||     |             || |||  : |
Db       62 APGGYPAPGGYPGAPQPGGAPSYPGVPPGQGFGVPPGGAGFSGYP-QPPSQSYGGGPAQV 120

Qy      206 ----------------------PSRDGT-------------------------ISEDTIR 218
                               ||:  |                          ::: |||
Db      121 PLPGGFPGGQMPSQYPGGQPTYPSQINTDSFSSYPVFSPVSLDYSSEPATVTQVTQGTIR 180

Qy      219 -ASLISAVSDK--LRWRMKEEMDGAQAELNALKRTEED---------------------- 253
             |:   |: |   ||  ||      || :: :     |
Db      181 PAANFDAIRDAEILRKAMKGFGTDEQAIVDVVANRSNDQRQKIKAAFKTSYGKDLIKDLK 240

Qy      254 ------------------------LKKGHQ--------KLEEMVTRLDQEVAEV----- 275
                                    |:|  |         :| : || :||: |:
Db      241 SELSGNMEELILALFMPPTYYDAWSLRKAMQGAGTQERVLIEILCTRTNQEIREIVRCYQ 300

Qy      276 -DKNIELLKKKDEELSSALEKM----------ENQSENNDIDEVIIPTAPLYKQILNLYA 324
              :  :| |    : |   |::          |||| |: :                  |
Db      301 SEFGRDLEKDIRSDTSGHFERLLVSMCQGNRDENQSINHQM-----------------A 342

Qy      325 EENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRKQFQLRALMQKARKTA 375
            :|:|    ::  ||    | :  |    :: | :|   |||| |:   : |
Db      343 QEDA--QRLYQAGE----GRLGTDESCFNMILATRSFPQLRATMEAYSRMA 387
```

RESULT 8
S29170
annexin VII - mouse
N;Alternate names: synexin
C;Species: Mus musculus (house mouse)
C;Date: 25-Feb-1994 #sequence_revision 01-Sep-1995 #text_change 13-Aug-1999
C;Accession: S29170; S46209; S51173
R;Zhang-Keck, Z.Y.; Burns, A.L.; Pollard, H.B.
Biochem. J. 289, 735-741, 1993
A;Title: Mouse synexin (annexin VII) polymorphisms and a phylogenetic comparison with ot
A;Reference number: S29170; MUID:93168121; PMID:7916616
A;Accession: S29170
A;Molecule type: mRNA
A;Residues: 1-463 <ZHA>
A;Cross-references: EMBL:L13129
R;Zhang-Keck, Z.Y.; Srivastava, M.; Kozak, C.A.; Caohuy, H.; Shirvan, A.; Burns, A.L.; P
Biochem. J. 301, 835-845, 1994
A;Title: Genomic organization and chromosomal localization of the mouse synexin gene.
A;Reference number: S46209; MUID:94330961; PMID:8053909
A;Accession: S46209
A;Molecule type: DNA
A;Residues: 1-463 <ZHF>
R;Zhang-Keck, Z.Y.; Burns, A.L.; Pollard, H.B.
submitted to the EMBL Data Library, May 1993
A;Reference number: S51173
A;Accession: S51173
A;Molecule type: mRNA
A;Residues: 1-144,'S',146-303,'A',305-463 <ZHW>
A;Cross-references: EMBL:L13129; NID:g293293; PIDN:AAA37238.1; PID:g293294
C;Genetics:
A;Gene: MGI:Anx7
A;Cross-references: MGI:88031

A;Map position: 14
A;Introns: 19/3; 87/1; 121/1; 142/3; 177/1; 208/3; 246/3; 303/3; 360/3; 386/1; 423/3
C;Superfamily: annexin VII; annexin repeat homology
C;Keywords: calcium channel; ion channel
F;163-234/Domain: annexin repeat homology <AX1>
F;235-306/Domain: annexin repeat homology <AX2>
F;318-390/Domain: annexin repeat homology <AX3>
F;394-463/Domain: annexin repeat homology <AX4>

Query Match 6.8%; Score 135.5; DB 2; Length 463;
Best Local Similarity 21.8%; Pred. No. 0.68;
Matches 78; Conservative 42; Mismatches 105; Indels 133; Gaps 17;

```
Qy      126 GEEP-PVFSRPTVSASYPPYTATGPPNTSYMPG-------------MPSGISAYPSGY- 169
            |: | |    |    :|||  : | |     |                 | |  ||| |
Db       30 GQYPYPSGFPPMGGGAYPPAPSGGYPGAGGYPAPGGYPAPGGYPGALSPGGPPAYPGGQG 89

Qy      170 ---PPNPSGYPGCPYPPA-----GP----------------YPATTSSQYPSQPPVTTV 204
               ||  :|: | | |||     ||                ||    : |||||   |
Db       90 FGAPPGGAGFSGYPQPPAQSYGGGPAQVPVPGGFPGGQMPSQYPG-GQAPYPSQPAAMTQ 148

Qy      205 GPS-----------RDGTISEDTIRA------SLISAVSDK----------------- 228
            |              ||  |   ::       :::  ||::
Db      149 GTQGTILPASNFDAMRDAEILRKAMKGFGTDEQAIVDVVSNRSNDQRQQIKAAFKTMYGK 208

Qy      229 -LRWRMKEEMDGAQAELNA---LKRTEED---LKKGHQ--------KLEEMVTRLDQEVA 273
             |   :| |: |   ||     :  |  |   |:|  |         :| : || :||:
Db      209 DLIKDLKSELSGNMEELILALFMPSTYYDAWSLRKAMQGAGTQERVLIEILCTRTNQEIR 268

Qy      274 EVDKNIELLKKKD------EELSSALEKM----------ENQSENNDIDEVIIPTAPLYK 317
            :: :  :|   :|       : |   |::          | || |: :
Db      269 DIVRCYQLEFGRDLEKDIRSDTSGHFERLLVSMCQGNRDERQSVNHQM----------- 316

Qy      318 QILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRKQFQLRALMQKARKTA 375
                |:|:|    ::  ||    | :  |    :: | :|   ||:| |:   : |
Db      317 ------AQEDA--QRLYQAGE----GRLGTDESCFNMILATRSFPQLKATMEAYSRMA 362
```

RESULT 9
T39608
zinc finger transcription factor - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 02-Sep-2000
C;Accession: T39608
R;Lyne, M.; Wood, V.; Rajandream, M.A.; Barrell, B.G.; Beck, A.; Reinhardt, R.
submitted to the EMBL Data Library, May 1998
A;Reference number: Z21866
A;Accession: T39608
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-827 <LYN>
A;Cross-references: EMBL:AL023554; PIDN:CAA19035.1; GSPDB:GN00067; SPDB:SPBC16G5.16
A;Experimental source: strain 972h-; cosmid c16G5
C;Genetics:
A;Gene: SPDB:SPBC16G5.16
A;Map position: 2
A;Introns: 30/3; 41/1
C;Superfamily: GAL4 zinc binuclear cluster homology
F;11-47/Domain: GAL4 zinc binuclear cluster homology <GL4>

Query Match 6.8%; Score 135.5; DB 2; Length 827;
Best Local Similarity 25.7%; Pred. No. 1.4;
Matches 63; Conservative 28; Mismatches 97; Indels 57; Gaps 14;

```
Qy        7 YRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSREL---VNLTGTIPVRYRGNIYNIP-- 61
            ||: |   |:      : :| :| :|  |   :  |    ::|| :|  :     |||
Db      588 YRESTYYSTI-----MEAIKNLLIAYDMNSSGTENLDATPDVTGQLPNNFSQRTSNIPRE 642

Qy       62 ---ICLWLLDT-YP--YNP------PICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKH 109
               ::  |  ||  |||       |  |  ||     :   :   ||  |  |     :
Db      643 FPQAQIFYSDAPYPGYYNPAQFQNAPTNFPMPTYGGRTQDQSYPRQNG--YPSYSDGNVY 700

Qy      110 PRSELLELIQIMIVIFGEEPPVFSRPTVSASYPPYTATGPP---NTSYMP------GMPS 160
```

```
Db        701 PHDRVM-------INYGS-----SMPTANGFYVPNTYSPVPFPYNTSYPPYMSPTSNMPQ 748

Qy        161 GISA---YPSGYPPNPSGYPGCPYPPAG----PYPATTSSQYPSQPPVTTVGPSRDGTIS 213
Db        749 AFQAYSQYPYQHPPFPLSEQMLPLPTSGVMMAPGAAKSGMPYPFIQP-----PSMTNQVA 803

Qy        214 EDTIR 218
Db        804 YPTVR 808
```

RESULT 10
D70509
hypothetical protein Rv1233c - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C;Accession: D70509
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: D70509
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-198 <COL>
A;Cross-references: GB:Z98260; GB:AL123456; NID:g3261826; PIDN:CAB10930.1; PID:e1299942;
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv1233c

Query Match 6.7%; Score 133; DB 2; Length 198;
Best Local Similarity 31.6%; Pred. No. 0.33;
Matches 50; Conservative 20; Mismatches 54; Indels 34; Gaps 10;

```
Qy        126 GEEPPVFSRP--------------TVSASYPP--YTATGPPNTSYMPGMPSGI------ 162
Db         17 GGPPPVGERPPEQPIADAPWAPPASSPMANHPPPAYPPSGYP-PAYQPGYPTGYPPPMPP 75

Qy        163 SAY-PSGYPP---NPSGYPGCPYPP-AGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTI 217
Db         76 GGYAPPGYPPPGTSSAGYGDIPYPPMPPPYGGSPGGYYPEPGYLDGYGPSQPGMNTMALV 135

Qy        218 RASLISAVSDKL--RWRMKEEMDGAQAELNALKRTEED 253
Db        136 --SLISALVGVLCCIGSIVGIVFGAIA-INQIKQTREE 170
```

RESULT 11
S72273
actin-depolymerizing protein N-WASP, brain - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 23-Apr-1998 #sequence_revision 01-May-1998 #text_change 05-Nov-1999
C;Accession: S72273
R;Miki, H.; Miura, K.; Takenawa, T.
EMBO J. 15, 5326-5335, 1996
A;Title: N-WASP, a novel actin-depolymerizing protein, regulates the cortical cytoskelet
A;Reference number: S72273; MUID:97050838; PMID:8895577
A;Accession: S72273
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-505 <MIK>
A;Cross-references: EMBL:D67066; NID:g1644231; PIDN:BAA11082.1; PID:d1011742; PID:g16442
A;Experimental source: brain

Query Match 6.7%; Score 133; DB 2; Length 505;
Best Local Similarity 27.0%; Pred. No. 1.1;
Matches 58; Conservative 31; Mismatches 66; Indels 60; Gaps 12;

```
Qy        126 GEEPPVFSR-PTVSASYPPYTATG------PPNTSYMPGMPSGISAYPSGYPPNP----- 173
Db        308 GAPPPPPSRAPTAAPPPPPSRPGVGAPPPPPNRMYPPPLPALPSSAPSGPPPPPPPLSV 367

Qy        174 SGYPGCPYPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRM 233
Db        368 SGSVAPPPPPPPPPP-------PGPPP--PPGLPSDGDHQVPTPAGSK-AALLDQIR--- 414

Qy        234 KEEMDGAQAELNALKRTEEDLK----KGHQKLEEMVTRLDQ--EVAEVDKN--------- 278
Db        415 ----EGAQ-----LKKVEQNSRPVSCSGRDALLDQIRQGIQLKSVTDAPESTPPAPAPTS 465

Qy        279 ------IELLKKKDEELSSALEKMENQSENNDIDE 307
Db        466 GIVGALMEVMQKRSKAIHSS-----DEDEDEDDDE 495
```

RESULT 12
T42731
atrophin-1 related protein - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 11-Jan-2000
C;Accession: T42731
R;Khan, F.A.; Margolis, R.L.; Loev, S.L.; Sharp, A.H.; Li, S.H.; Ross, C.A.
submitted to the EMBL Data Library, December 1995
A;Description: CDNA sequence and expression of an atrophin-1 (DRPLA disease gene) rel
A;Reference number: Z22250
A;Accession: T42731
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1006 <KHA>
A;Cross-references: EMBL:U44091; NID:g1297310; PID:g1209103; PIDN:AAA98970.1
C;Genetics:
A;Gene: ARP

Query Match 6.5%; Score 130; DB 2; Length 1006;
Best Local Similarity 20.0%; Pred. No. 3.8;
Matches 64; Conservative 41; Mismatches 111; Indels 104; Gaps 11;

```
Qy         44 NLTGTIPVRYRGNIYNIPICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPY 103
Db        388 HLSGPSPFSMNANL----------------PPPPALKPLSSLST---------------- 415

Qy        104 LHDWKHPRSELLELIQIMIVIFGEEPPVFSRP------TVSASYPPYTATGP-------- 149
Db        416 ----HHPPSAHPPPLQLM----PQSQPLPSSPAQPPGLTQSQSLPPPAASHPTTGGLHQV 467

Qy        150 ------PNTSYMPGMPSGI---SAYPSGYPP------------------NPSGYPGCPY 181
Db        468 PSQSPFPQHPFVPGGPPPITPPSCPPTSTPPAGPSSSSQPPCSAAVSSGGNVPGAPSCPL 527

Qy        182 PPAG-PYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDG- 239
Db        528 PAVQIKEEALDEAEEPESPPPPPRSPSPEPTVVDTPSHASQSA--------RFYKHLDRG 579

Qy        240 ----AQAEL--------NALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDE 287
Db        580 YNSCARTDLYFMPLAGSKLAKKREEAIEKAKREAEQKAREERERERERERERERERERER 639

Qy        288 ELSSALEKMENQSENNDIDE 307
Db        640 EAERAAQKASSSAHEGRLSD 659
```

RESULT 13
T28754
hypothetical protein T08B2.5 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C;Accession: T28754
R;Blanchard, M.; Wamsley, P.
submitted to the EMBL Data Library, May 1997
A;Description: The sequence of C. elegans cosmid T08B2.
A;Reference number: Z20518

```
A;Accession: T28754
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-669 <BLA>
A;Cross-references: EMBL:AF000263; PIDN:AAC48191.1; GSPDB:GN00019; CESP:T08B2.5
A;Experimental source: strain Bristol N2; clone T08B2
C;Genetics:
A;Gene: CESP:T08B2.5
A;Map position: 1
A;Introns: 33/3; 68/3; 108/1; 177/1; 216/1; 242/3; 297/1; 418/1; 511/3; 623/1

  Query Match             6.5%;  Score 129;  DB 2;  Length 669;
  Best Local Similarity   20.7%;  Pred. No. 2.6;
  Matches   81;  Conservative   60;  Mismatches  193;  Indels   58;  Gaps   14;

Qy         39 SRELVNLTGTIPVRYRGNIYNIPICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDAN-- 96
              :: |:|||   |:| :     :  |   :                    :: ||     |
Db        158 AKMLLNLTYKSPIRIKDKDVQVSWCRDSMSKLIQQQMTMLTAGPGKNSVATGSLAQGNMT 217

Qy         97 -GKIYLPYLHDWKHPRSELLELIQIMIVIFGEE----PPVFSRPTVSASY-PPYTATGPP 150
               :|    |      |    :: |||  :        || || |  : |  ||  || |
Db        218 GAEIAAAALSKANAVRQASQQVGQIMTGLPAPAANMIPPNFSVPPPNLSVPPPMQATQPE 277

Qy        151 NTSYMPGM---PSGISAYPSGYPPNPSGYPGCP-------------YPPA-GPYPATTSS 193
              : : : ||   | |:   |   ||||  :   |               | | | |    :|
Db        278 HQNGVIGMTQTPKGL--LPRYLPPNPLTFAHDPNLGYYVDPITKFCYDSATGYYFNNATS 335

Qy        194 Q-----------YPSQPP--VTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGA 240
              |           :| : |  : :  |  :     :    :   | |  :|  ::| | |
Db        336 QWCNWDLTHHTYFPVETPAVINSADPEEEKMKKNEEGPKTAQDLVKDMAKWAKRQEKDKK 395

Qy        241 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 300
              : :::   :  : ::  :    |   |: |  |  | :    :::|:|      : |  |
Db        396 KVQISIKGKETKGIELKNVFSNEKQKRIAQTAALFDDD----EEEDQEEEVRGRRSEEPS 451

Qy        301 ENNDIDEVIIPTAPLYKQILNLY-AEENAIEDTIFYLGEALRRGVIDLDVFLKHVRL--L 357
               ::    |  | ||    : ::  | | |: |         :|   |::|  |||    |
Db        452 TSSSFSSV--PLAPRKSTLQDVRDAMERALYDETKKTCMLCKRAFSDVEVLRKHVEKSDL 509

Qy        358 SRKQFQLR---------ALMQKARKTAGLSDL 380
              |   : :         | :|:  : |   ||
Db        510 HRNNLEAKRAEWGRETAAKLQEEEEDASAPDL 541


RESULT 14
A53152
annexin XI - human
C;Species: Homo sapiens (man)
C;Date: 07-Jul-1995 #sequence_revision 07-Jul-1995 #text_change 13-Aug-1999
C;Accession: A53152
R;Misaki, Y.; Pruijn, G.J.M.; van der Kemp, A.W.C.M.; van Venrooij, W.J.
J. Biol. Chem. 269, 4240-4246, 1994
A;Title: The 56K autoantigen is identical to human annexin XI.
A;Reference number: A53152; MUID:94140847; PMID:7508441
A;Accession: A53152
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-505 <MIS>
A;Cross-references: GB:L19605; NID:g457128; PIDN:AAA19734.1; PID:g457129
C;Genetics:
A;Gene: GDB:ANX11
A;Cross-references: GDB:313076
A;Map position: 9q11-9q22
C;Superfamily: annexin VII; annexin repeat homology
C;Keywords: calcium binding; duplication; endonexin fold; glycoprotein; phospholipid bin
F;203-274/Domain: annexin repeat homology <AX1>
F;275-346/Domain: annexin repeat homology <AX2>
F;358-430/Domain: annexin repeat homology <AX3>
F;434-505/Domain: annexin repeat homology <AX4>

  Query Match             6.4%;  Score 128;  DB 2;  Length 505;
  Best Local Similarity   28.1%;  Pred. No. 2.1;
  Matches   55;  Conservative   15;  Mismatches   76;  Indels   50;  Gaps    9;

Qy        139 ASYPPYT--ATGPPNTSYMPGMPSGISAYPSGYPPN--PS--GYPGCPYP---------- 182
              | |||      | | ::  |  | |:   | | ||:  ||   ||| | |
Db         80 AGYPPVPPGGFGQPPSAQQPVPPYGMYPPPGGNPPSRMPSYPPYPGAPVPGQPMPPPGQQ 139

Qy        183 PAGPYPATTSSQYPSQPPVTTVG-----PSRDGTISEDTI----------RASLISAV- 225
              | | ||     || ||||   |      ||  |     |:           | ::  |
Db        140 PPGAYPGQPPVTYPGQPPVPLPGQQQPVPSYPGYPGSGTVTPAVPPTQFGSRGTITDAPG 199

Qy        226 ------SDKLRWRMKEEMDGAQAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNI 279
                   :: ||  ||      || :: |       |:  | |    |   :
Db        200 FDPLRDAEVLRKAMKGFGTDEQAIIDCLGSRSN--KQRQQILLSFKTAYGK--------- 248

Qy        280 ELLKKKDEELSSALEK 295
              :|:|    |||   ||
Db        249 DLIKDLKSELSGNFEK 264


RESULT 15
JC5368
dynactin 1 - mouse
N;Alternate names: p150 Glued
C;Species: Mus musculus (house mouse)
C;Date: 28-May-1997 #sequence_revision 18-Jul-1997 #text_change 05-Nov-1999
C;Accession: JC5368
R;Jang, W.; Weber, J.S.; Tokito, M.K.; Holzbaur, E.L.F.; Meisler, M.H.
Biochem. Biophys. Res. Commun. 231, 344-347, 1997
A;Title: Mouse p150Glued (dynactin 1) cDNA sequence and evaluation as a candidate for
A;Reference number: JC5368; MUID:97223454; PMID:9070275
A;Accession: JC5368
A;Molecule type: mRNA
A;Residues: 1-1281 <JAN>
A;Cross-references: GB:U60312; NID:g2104494; PIDN:AAB57773.1; PID:g2104495
A;Experimental source: brain
C;Comment: This protein is a member of the oligomeric dynactin complex that is requir
C;Genetics:
A;Gene: Dctn1
A;Map position: 6

  Query Match             6.4%;  Score 127.5;  DB 2;  Length 1281;
  Best Local Similarity   24.5%;  Pred. No. 7.3;
  Matches   69;  Conservative   46;  Mismatches  104;  Indels   63;  Gaps   14;

Qy        134 RPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCPYPPAGPYPATTSS 193
              :||  ||    |    |::|  |   :      |  |  |:  |     |  | || ||
Db        151 KPTRPAS----TGVAGPSSSLGPSGSASAGELSSSEPSTPAQTPLA--APIIPTPALTSP 204

Qy        194 QYPSQPPVTTVGPSRDGTIS-----EDTIRASLISAVSDKLRWR--------------MK 234
                : ||: :     :|  :     |: :    :    || : :               |
Db        205 --GAAPPLPSPSKEEEGLRAQVRDLEEKLETLRLKRSEDKAKLKELEKHKIQLEQVQEWK 262

Qy        235 EEMDGAQAEL-----NALKRTEEDLKKGHQKLEEM--------VTRLDQEVAE-----VD 276
               :|   ||:|      | |  :| |:   : :|||         :  ||:|:||     :
Db        263 SKMQEQQADLQRRLKEARKEAKEALEAKERYMEEMADTADAIEMATLDKEMAEERAESLQ 322

Qy        277 KNIELLKKKDEELSSALEKMENQSENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYL 336
              : :| ||:: :||:: || :: : |    |      |  |:  |    |:||     |
Db        323 QEVEALKERVDELTTDLEILKAEIEEKGSD----GAASSYQ--LKQLEEQNA------RL 370

Qy        337 GEALRRGVIDLDVFLK--HV---RLLSRKQFQLRALMQKARK 373
              :|| | : ||    |  ||   :|: :|  :|  : |:  :
Db        371 KDALVR-MRDLSSSEKQEHVKLQKLMEKKNQELEVVRQQRER 411


Search completed: June  3, 2003, 14:09:56
Job time : 29.0342 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: June 3, 2003, 14:05:55 ; Search time 14.0184 Seconds
(without alignments)
1127.267 Million cell updates/sec

Title: US-09-804-690-2
Perfect score: 1996
Sequence: 1 MMSKYKYRDLTVRQTVNVIA.....FQLRALMQKARKTAGLSDLY 381

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1996	100.0	391	1	T101_MOUSE	Q61187 mus musculu
2	1900.5	95.2	390	1	T101_HUMAN	Q99816 homo sapien
3	150	7.5	466	1	ANX7_HUMAN	P20073 homo sapien
4	147	7.4	503	1	ANXB_MOUSE	P97384 mus musculu
5	142	7.1	212	1	YCA8_YEAST	P25604 saccharomyc
6	137.5	6.9	503	1	ANXB_RABIT	P33477 oryctolagus
7	135.5	6.8	463	1	ANX7_MOUSE	Q07076 mus musculu
8	133	6.7	505	1	WASL_BOVIN	Q95107 bos taurus
9	131	6.6	1280	1	DYNA_RAT	P28023 rattus norv
10	128	6.4	505	1	ANXB_HUMAN	P50995 homo sapien
11	127.5	6.4	1281	1	DYNA_MOUSE	O08788 mus musculu
12	127	6.4	1224	1	DYNA_CHICK	P35458 gallus gall
13	126.5	6.3	505	1	WASL_HUMAN	O00401 homo sapien
14	123.5	6.2	376	1	VASP_MOUSE	P70460 mus musculu
15	122.5	6.1	501	1	WASL_RAT	O08816 rattus norv
16	122.5	6.1	503	1	ANXB_BOVIN	P27214 bos taurus
17	122	6.1	746	1	PCAP_HUMAN	Q96rn5 homo sapien
18	122	6.1	978	1	RA50_AQUAE	O67124 aquifex aeo
19	120.5	6.0	338	1	FOSB_HUMAN	P53539 homo sapien
20	119.5	6.0	753	1	ZIN_HUMAN	Q9nr13 homo sapien
21	119.5	6.0	815	1	MK07_HUMAN	Q13164 homo sapien
22	119.5	6.0	1733	1	RPB1_YEAST	P04050 saccharomyc
23	119	6.0	380	1	VASP_HUMAN	P50552 homo sapien
24	119	6.0	792	1	PCAP_MOUSE	Q924h2 mus musculu
25	118.5	5.9	384	1	VASP_CANFA	P50551 canis famil
26	118.5	5.9	559	1	WAS1_HUMAN	Q92558 homo sapien
27	118.5	5.9	760	1	ZIN_MOUSE	P58404 mus musculu
28	118.5	5.9	1647	1	SN24_HUMAN	P51532 homo sapien
29	117.5	5.9	261	1	LEG3_RAT	P08699 rattus norv
30	117.5	5.9	1054	1	S24A_ARATH	Q9sfu0 arabidopsis
31	117.5	5.9	1185	1	DRPL_HUMAN	P54259 homo sapien
32	117.5	5.9	1234	1	PIP3_MOUSE	P51432 mus musculu
33	117	5.9	609	1	YSW1_YEAST	P38280 saccharomyc
34	117	5.9	964	1	YQY1_CAEEL	Q09560 caenorhabdi
35	117	5.9	1453	1	CA11_MOUSE	P11087 mus musculu
36	117	5.9	1752	1	RPB1_SCHPO	P36594 schizosacch
37	116.5	5.8	938	1	CAFA_HUMAN	Q13111 homo sapien
38	116.5	5.8	1278	1	DYNA_HUMAN	Q14203 homo sapien
39	116.5	5.8	1887	1	RPB1_DROME	P04052 drosophila
40	116	5.8	338	1	FOSB_MOUSE	P13346 mus musculu
41	116	5.8	1550	1	ACS1_ACEXY	P21877 acetobacter
42	115.5	5.8	467	1	CBPA_DICDI	P35085 dictyosteli
43	115.5	5.8	1790	1	USO1_YEAST	P25386 saccharomyc
44	115	5.8	605	1	VP40_EBV	P03234 epstein-bar
45	115	5.8	751	1	FP1_MYTGA	Q27409 mytilus gal

ALIGNMENTS

RESULT 1
T101_MOUSE
ID T101_MOUSE STANDARD; PRT; 391 AA.
AC Q61187;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tumor susceptibility gene 101 protein.
GN TSG101.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fibroblast;
RX MEDLINE=96201522; PubMed=8616888;
RA Li L., Cohen S.N.;
RT "Tsg101: a novel tumor susceptibility gene isolated by controlled
RT homozygous functional knockout of allelic loci in mammalian cells.";
RL Cell 85:319-329(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SvJ; TISSUE=Mammary gland;
RX MEDLINE=99054675; PubMed=9840940;
RA Wagner K.-U., Dierisseau P., Rucker E.B. III, Robinson G.W.,
RA Hennighausen L.;
RT "Genomic architecture and transcriptional activation of the mouse and
RT human tumor susceptibility gene TSG101: common types of shorter
RT transcripts are true alternative splice variants.";
RL Oncogene 17:2761-2770(1998).
RN [3]
RP INTERACTION WITH DMAP1.
RX MEDLINE=20347709; PubMed=10888872;
RA Rountree M.R., Bachman K.E., Baylin S.B.;
RT "DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at
RT replication foci.";
RL Nat. Genet. 25:269-277(2000).
CC -!- FUNCTION: MAY BE INVOLVED IN CELL GROWTH AND DIFFERENTIATION AND
CC ACT AS A NEGATIVE GROWTH REGULATOR.
CC -!- SUBUNIT: INTERACTS WITH STATHMIN AND DMAP1.
CC -!- SUBCELLULAR LOCATION: MAINLY CYTOPLASMIC. DEPENDING ON THE STAGE
CC OF THE CELL CYCLE, DETECTED IN THE NUCLEUS.
CC -!- TISSUE SPECIFICITY: UBIQUITOUS. HIGHER EXPRESSION IN BRAIN AND
CC MAMMARY GLAND. LOWER EXPRESSION IN LIVER AND TUMORAL TISSUES.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED AT ALL STAGES OF MAMMARY GLAND
CC DEVELOPMENT, BUT AT LOWER RATE AT EARLY AND MID PREGNANCY.
CC EXPRESSED IN 1-CELL AND 2-CELL STAGE EMBRYOS.
CC -!- DOMAIN: THE COILED COIL DOMAIN MAY INTERACT WITH STATHMIN.
CC ----------

```
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; U52945; AAC53586.1; -.
DR   EMBL; AF060868; AAC83576.1; -.
DR   MGD; MGI:106581; tsg101.
DR   InterPro; IPR000608; UBQ_conjugat.
DR   SMART; SM00212; UBCc; 1.
KW   Growth regulation; Coiled coil.
FT   DOMAIN      237    317       COILED COIL (POTENTIAL).
SQ   SEQUENCE   391 AA;  44123 MW;  79558EC535093492 CRC64;

  Query Match             100.0%;  Score 1996;  DB 1;  Length 391;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-106;
  Matches  381;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         11 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNI 70

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQI 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         71 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQI 130

Qy        121 MIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCP 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        131 MIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCP 190

Qy        181 YPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        191 YPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGA 250

Qy        241 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        251 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 310

Qy        301 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        311 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 370

Qy        361 QFQLRALMQKARKTAGLSDLY 381
              |||||||||||||||||||||
Db        371 QFQLRALMQKARKTAGLSDLY 391


RESULT 2
T101_HUMAN
ID   T101_HUMAN     STANDARD;      PRT;   390 AA.
AC   Q99816;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Tumor susceptibility gene 101 protein.
GN   TSG101.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=97148696; PubMed=9019400;
RA   Li L., Li X., Francke U., Cohen S.N.;
RT   "The TSG101 tumor susceptibility gene is located in chromosome 11 band
RT   p15 and is mutated in human breast cancer.";
RL   Cell 88:143-154(1997).
RN   [2]
RP   ERRATUM.
RX   MEDLINE=99080983; PubMed=9867424;
RA   Li L., Francke U., Cohen S.N.;
RL   Cell 93:660-660(1998).
RN   [3]
RP   ALTERNATIVE SPLICING (ISOFORM 2).
RX   MEDLINE=98031856; PubMed=9366528;
RA   Gayther S.A., Barski P., Batley S.J., Li L., de Foy K.A., Cohen S.N.,
RA   Ponder B.A., Caldas C.;
RT   "Aberrant splicing of the TSG101 and FHIT genes occurs frequently in
RT   multiple malignancies and in normal tissues and mimics alterations
RT   previously described in tumours.";
RL   Oncogene 15:2119-2126(1997).
RN   [4]
RP   ALTERNATIVE SPLICING.
RX   MEDLINE=97384943; PubMed=9242438;
RA   Lee M.P., Feinberg A.P.;
RT   "Aberrant splicing but not mutations of TSG101 in human breast
RT   cancer.";
RL   Cancer Res. 57:3131-3134(1997).
RN   [5]
RP   ALTERNATIVE SPLICING.
RX   MEDLINE=99054675; PubMed=9840940;
RA   Wagner K.-U., Dierisseau P., Rucker E.B. III, Robinson G.W.,
RA   Hennighausen L.;
RT   "Genomic architecture and transcriptional activation of the mouse and
RT   human tumor susceptibility gene TSG101: common types of shorter
RT   transcripts are true alternative splice variants.";
RL   Oncogene 17:2761-2770(1998).
RN   [6]
RP   INTERACTION WITH DMAP1.
RX   MEDLINE=20347709; PubMed=10888872;
RA   Rountree M.R., Bachman K.E., Baylin S.B.;
RT   "DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at
RT   replication foci.";
RL   Nat. Genet. 25:269-277(2000).
CC   -!- FUNCTION: MAY BE INVOLVED IN CELL GROWTH AND DIFFERENTIATION AND
CC       ACT AS A NEGATIVE GROWTH REGULATOR.
CC   -!- SUBUNIT: INTERACTS WITH STATHMIN AND DMAP1.
CC   -!- SUBCELLULAR LOCATION: MAINLY CYTOPLASMIC. DEPENDING ON THE STAGE
CC       OF THE CELL CYCLE, DETECTED IN THE NUCLEUS.
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; 1 (SHOWN HERE) AND 2;
CC       ARE PRODUCED BY ALTERNATIVE SPLICING. ISOFORM 2 IS DETECTED IN
CC       NORMAL AS WELL AS CANCER TISSUES. SEVERAL SHORTER ISOFORMS ARE
CC       DETECTED IN PRIMARY BREAST CANCERS AND OTHER TUMORS.
CC   -!- TISSUE SPECIFICITY: HEART, BRAIN, PLACENTA, LUNG, LIVER, SKELETAL,
CC       KIDNEY AND PANCREAS.
CC   -!- DOMAIN: THE COILED COIL DOMAIN MAY INTERACT WITH STATHMIN.
CC   ------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; U82130; AAC52083.1; -.
DR   Genew; HGNC:15971; TSG101.
DR   MIM; 601387; -.
DR   InterPro; IPR000608; UBQ_conjugat.
DR   SMART; SM00212; UBCc; 1.
KW   Growth regulation; Coiled coil; Alternative splicing.
FT   DOMAIN      235    316       COILED COIL (POTENTIAL).
FT   VARSPLIC     15    119       MISSING (IN ISOFORM 2).
SQ   SEQUENCE   390 AA;  43944 MW;  ADD6912FC22DF162 CRC64;

  Query Match             95.2%;  Score 1900.5;  DB 1;  Length 390;
  Best Local Similarity   94.5%;  Pred. No. 5.4e-101;
  Matches  360;  Conservative   11;  Mismatches    9;  Indels    1;  Gaps    1;

Qy          1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNI 60
              |:||||||||||:||||| :|||||||||||||||||||:|||||||| |||| |||
Db         11 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 70

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQI 120
              ||||||||||||||||||||||||||||||||||||||||||||||||:|||:|:|| |||:
Db         71 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 130

Qy        121 MIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCP 180
```

```
                ||||:||:|||||||| :|||||||  |||||||||||||| ||| ||||||||||||||||
Db        131 MIVVFGDEPPVFSRP-ISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCP 189

Qy        181 YPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGA 240
              ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        190 YPPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRA 249

Qy        241 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        250 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 309

Qy        301 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        310 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 369

Qy        361 QFQLRALMQKARKTAGLSDLY 381
              |||||||||||||||||||||
Db        370 QFQLRALMQKARKTAGLSDLY 390
```

RESULT 3
ANX7_HUMAN
ID ANX7_HUMAN STANDARD; PRT; 466 AA.
AC P20073;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Annexin A7 (Annexin VII) (Synexin).
GN ANXA7 OR ANX7.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=89264510; PubMed=2542947;
RA Burns A.L., Magendzo K., Shirvan A., Srivastava M., Rojas E.,
RA Alijani M.R., Pollard H.B.;
RT "Calcium channel activity of purified human synexin and structure of
RT the human synexin gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:3798-3802(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: CALCIUM/PHOSPHOLIPID-BINDING PROTEIN WHICH PROMOTES
CC   MEMBRANE FUSION AND IS INVOLVED IN EXOCYTOSIS.
CC -!- DOMAIN: A pair of annexin repeats may form one binding site for
CC   calcium and phospholipid.
CC -!- SIMILARITY: BELONGS TO THE ANNEXIN FAMILY.
CC -!- SIMILARITY: CONTAINS 4 ANNEXIN REPEATS.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; J04543; AAA36616.1; -.
DR EMBL; BC002632; AAH02632.1; -.
DR PIR; A32554; LUHU7.
DR HSSP; P26256; 1DM5.
DR Genew; HGNC:545; ANXA7.
DR MIM; 186360; -.
DR InterPro; IPR001464; Annexin.
DR Pfam; PF00191; annexin; 4.
DR PRINTS; PR00196; ANNEXIN.
DR ProDom; PD000143; Annexin; 4.
DR SMART; SM00335; ANX; 4.
DR PROSITE; PS00223; ANNEXIN; 4.

KW Annexin; Calcium/phospholipid-binding; Repeat.
FT DOMAIN 1 143 REPEAT-RICH REGION.
FT REPEAT 172 232 ANNEXIN 1.
FT REPEAT 244 304 ANNEXIN 2.
FT REPEAT 327 387 ANNEXIN 3.
FT REPEAT 403 463 ANNEXIN 4.
FT DOMAIN 5 20 3 X 5 AA TANDEM REPEATS OF G-Y-P-P-X.
FT REPEAT 5 9 1.
FT REPEAT 10 14 2.
FT REPEAT 16 20 3.
SQ SEQUENCE 466 AA; 50315 MW; 09A6760729D45FCD CRC64;

  Query Match 7.5%; Score 150; DB 1; Length 466;
  Best Local Similarity 21.3%; Pred. No. 0.086;
  Matches 83; Conservative 36; Mismatches 92; Indels 178; Gaps 16;

```
Qy        140 SYPPYTATG-PPNTSYMPG------MPSGISAYPSGYP---------------------- 170
              ||| |  || ||   | |        |||   ||||:|
Db          2 SYPGYPPTGYPPFPGYPPAGQESSFPPSGQYPYPSGFPPMGGGAYPQVPSSGYPGAGGYP 61

Qy        171 -----PNPSGYPGCPYPPAGP-YPATTSSQ-------------YPSQPPVTTVG------ 205
                  | | |||| | |   | ||     |              || |||  : |
Db         62 APGGYPAPGGYPGAPQPGGAPSYPGVPPGQGFGVPPGGAGFSGYP-QPPSQSYGGGPAQV 120

Qy        206 ----------------------PSRDGTISEDT--------------------------- 216
                                   ||:  |::: |
Db        121 PLPGGFPGGQMPSQYPGGQPTYPSQPATVTQVTQGTIRPAANFDAIRDAEILRKAMKGFG 180

Qy        217 --------------------IRASLISAVSDKLRWRMKEEMDGAQAELNA---LKRTEED 253
                               |:|:  ::    |   :| |: |   ||     :  | |
Db        181 TDEQAIVDVVANRSNDQRQKIKAAFKTSYGKDLIKDLKSELSGNMEELILALFMPPTYYD 240

Qy        254 ---LKKGHQ--------KLEEMVTRLDQEVAEV------DKNIELLKKKDEELSSALEKM 296
                |:|  |         :| : || :||: |:      :   :| |    : |   |::
Db        241 AWSLRKAMQGAGTQERVLIEILCTRTNQEIREIVRCYQSEFGRDLEKDIRSDTSGHFERL 300

Qy        297 ----------ENQSENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVID 346
                       |||| |: :                 |:|:|    ::  ||     | :  .
Db        301 LVSMCQGNRDENQSINHQM----------------AQEDA--QRLYQAGE----GRLG 336

Qy        347 LDVFLKHVRLLSRKQFQLRALMQKARKTA 375
              |    :: | :|   |||| |:   : |
Db        337 TDESCFNMILATRSFPQLRATMEAYSRMA 365
```

RESULT 4
ANXB_MOUSE
ID ANXB_MOUSE STANDARD; PRT; 503 AA.
AC P97384;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Annexin A11 (Annexin XI) (Calcyclin-associated annexin 50) (CAP-50).
GN ANXA11 OR ANX11.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97092887; PubMed=8938449;
RA Fernandez M.P., Jenkins N.A., Gilbert D.J., Copeland N.G.,
RA Morgan R.O.;
RT "Sequence and chromosomal localization of mouse annexin XI.";
RL Genomics 37:366-374(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SvJ;
RX MEDLINE=20469408; PubMed=11013079;
RA Bances P., Fernandez M.R., Rodriguez-Garcia M.I., Morgan R.O.,
RA Fernandez M.P.;
RT "Annexin A11 (ANXA11) gene structure as the progenitor of paralogous
RT annexins and source of orthologous cDNA isoforms.";

```
RL   Genomics 69:95-103(2000).
CC   -!- FUNCTION: BINDS SPECIFICALLY TO CALCYCLIN IN A CALCIUM-DEPENDENT
CC       MANNER.
CC   -!- DOMAIN: A pair of annexin repeats may form one binding site for
CC       calcium and phospholipid.
CC   -!- SIMILARITY: BELONGS TO THE ANNEXIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 4 ANNEXIN REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U65986; AAB42012.1; -.
DR   EMBL; AJ289760; CAB94770.1; -.
DR   EMBL; AJ289761; CAB94770.1; JOINED.
DR   EMBL; AJ289762; CAB94770.1; JOINED.
DR   EMBL; AJ289763; CAB94770.1; JOINED.
DR   EMBL; AJ289764; CAB94770.1; JOINED.
DR   EMBL; AJ289765; CAB94770.1; JOINED.
DR   EMBL; AJ289766; CAB94770.1; JOINED.
DR   EMBL; AJ289767; CAB94770.1; JOINED.
DR   EMBL; AJ289768; CAB94770.1; JOINED.
DR   EMBL; AJ289769; CAB94770.1; JOINED.
DR   HSSP; P13214; 1ANN.
DR   SWISS-2DPAGE; P97384; MOUSE.
DR   MGD; MGI:108481; Anxa11.
DR   InterPro; IPR001464; Annexin.
DR   Pfam; PF00191; annexin; 4.
DR   PRINTS; PR00196; ANNEXIN.
DR   ProDom; PD000143; Annexin; 4.
DR   SMART; SM00335; ANX; 4.
DR   PROSITE; PS00223; ANNEXIN; 4.
KW   Annexin; Calcium/phospholipid-binding; Repeat.
FT   REPEAT      207    267       ANNEXIN 1.
FT   REPEAT      279    339       ANNEXIN 2.
FT   REPEAT      363    423       ANNEXIN 3.
FT   REPEAT      438    498       ANNEXIN 4.
SQ   SEQUENCE   503 AA;  54111 MW;  424B1345E0F4EC8A CRC64;

  Query Match             7.4%;  Score 147;  DB 1;  Length 503;
  Best Local Similarity   26.3%;  Pred. No. 0.14;
  Matches   77;  Conservative   29;  Mismatches   95;  Indels   92;  Gaps   16;

Qy        125 FGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCPYPPA 184
              ||: ||   :|      |||    ||  :  |||||    ||: ||  |   || | ||
Db         89 FGQPPPA-QQPV-----PPYGMYPPPGGNPPPGMPS----YPA-YPGAP--VPGQPMPPT 135

Qy        185 GPYPATTSSQYPSQPPVTTVG-----------PSRDGTISEDTIRASL----------I 222
              |  |      || |||:|  |            ||  |     ||  ::  ~        |
Db        136 GQQP---PGAYPGQPPMTYPGQSPMPPPGQQPVPSYPGYSGSSTITPAVPPAQFGNRGTI 192

Qy        223 SAVS--------DKLRWRMKEEMDGAQAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAE 274
              :| |        : ||  ||      || :: |       |:  | |    |   :
Db        193 TAASGFDPLRDAEVLRKAMKGFGTDEQAIIDCLGSRSN--KQRQQILLSFKTAYGK---- 246

Qy        275 VDKNIELLKKKDEELSSALEKMENQSENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIF 334
                  :|:|    |||   |                    | || |   :  :   ::
Db        247 -----DLIKDLKSELSGNFE--------------------KTILALM--KTPVLFDVY 277

Qy        335 YLGEALRRGVID----LDVFL----KHVRLLSRKQFQLRALMQKARKTAGLSD 379
                : ||::    |    :::|     :|:| |||     :   ||  : |  ||
Db        278 EIKEAIKGAGTDEACLIEIFASRSNEHIRELSR---AYKTEFQKTLEEAIRSD 327


RESULT 5
YCA8_YEAST
ID   YCA8_YEAST     STANDARD;      PRT;   212 AA.
AC   P25604; P87010;
DT   01-MAY-1992 (Rel. 22, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical 24.1 kDa protein in ILV6-CWH36 intergenic region.
GN   YCL008C OR YCL8C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Oliver S.G., Anwar R., Brown A., Gent M.E., Indge K.J., James C.M.,
RA   Stateva L.I.;
RL   Submitted (MAR-1992) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   REVISIONS.
RA   Gromadka R.;
RL   Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X59720; CAA42351.1; -.
DR   PIR; S19410; S19410.
DR   SGD; S0000514; STP22.
DR   InterPro; IPR000608; UBQ_conjugat.
DR   SMART; SM00212; UBCc; 1.
DR   PROSITE; PS00183; UBIQUITIN_CONJUGAT_1; FALSE_NEG.
DR   PROSITE; PS50127; UBIQUITIN_CONJUGAT_2; 1.
KW   Hypothetical protein; Ubl conjugation pathway; Ligase.
SQ   SEQUENCE   212 AA;  24086 MW;  AE5309BB58A2B2B8 CRC64;

  Query Match             7.1%;  Score 142;  DB 1;  Length 212;
  Best Local Similarity   23.7%;  Pred. No. 0.095;
  Matches   57;  Conservative   37;  Mismatches   99;  Indels   48;  Gaps   11;

Qy         64 LWLLDTYPYNPPIC------FVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLEL 117
              :|:   ||  ||         |   | | ::   :::|:|| | || || |      |: :
Db          1 MWVPSMYPVKPPFISINLENFDMNTISSSLPIQEYIDSNGWIALPILHCWDPAAMNLIMV 60

Qy        118 IQIMIVIFGEEPPVFSRPTVSASYPPYTAT---GPPNTSYMPGMPSGISAYPSGYPPNPS 174
              :| :: :   |||    |    | ||   |      ||  :|  |      |   || |
Db         61 VQELMSLL-HEPPQDQAP----SLPPKPNTQLQQEQNTPPLPPKPKS----PHLKPPLP- 110

Qy        175 GYPGCPYPPAGPYPATTSSQYPSQPPVTTVGPSRD-------GTISEDTIRASLISAVSD 227
                       ||  | ||: :          | : |:          |:  :  |   :  |:|
Db        111 -------PPPPPQPASNALDLMDMDN-TDISPTNHHEMLQNLQTVVNELYRED-VDYVAD 161

Qy        228 KLRWR---MKEEMDGAQAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKK 284
              |:  |   |:|          :: |  | :      |  :   ::| :  ::  |::| :
Db        162 KILTRQTVMQE---------SIARFHEIIAIDKNHLRAVEQAIEQTMHSLNAQIDVLNR 211

Qy        285 K 285
              |
Db        212 K 212


RESULT 6
ANXB_RABIT
ID   ANXB_RABIT     STANDARD;      PRT;   503 AA.
AC   P33477;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Annexin A11 (Annexin XI) (Calcyclin-associated annexin 50) (CAP-50).
GN   ANXA11 OR ANX11.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
```

```
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Lung;
RX   MEDLINE=92378579; PubMed=1380798;
RA   Tokumitsu H., Mizutani A., Muramatsu M.-A., Yokota T., Arai K.-I.,
RA   Hidaka H.;
RT   "Molecular cloning of rabbit CAP-50, a calcyclin-associated annexin
RT   protein.";
RL   Biochem. Biophys. Res. Commun. 186:1227-1235(1992).
CC   -!- FUNCTION: BINDS SPECIFICALLY TO CALCYCLIN IN A CALCIUM-DEPENDENT
CC       MANNER.
CC   -!- DOMAIN: A pair of annexin repeats may form one binding site for
CC       calcium and phospholipid.
CC   -!- SIMILARITY: BELONGS TO THE ANNEXIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 4 ANNEXIN REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D10883; BAA01705.1; -.
DR   PIR; JH0694; LURB11.
DR   HSSP; P13214; 1ANN.
DR   InterPro; IPR001464; Annexin.
DR   Pfam; PF00191; annexin; 4.
DR   PRINTS; PR00196; ANNEXIN.
DR   ProDom; PD000143; Annexin; 4.
DR   SMART; SM00335; ANX; 4.
DR   PROSITE; PS00223; ANNEXIN; 4.
KW   Annexin; Calcium/phospholipid-binding; Repeat.
FT   REPEAT      207    267       ANNEXIN 1.
FT   REPEAT      279    339       ANNEXIN 2.
FT   REPEAT      363    423       ANNEXIN 3.
FT   REPEAT      438    498       ANNEXIN 4.
SQ   SEQUENCE   503 AA;  54034 MW;  44C15F290770AC9F CRC64;

  Query Match             6.9%;  Score 137.5;  DB 1;  Length 503;
  Best Local Similarity   22.7%;  Pred. No. 0.48;
  Matches   75;  Conservative   42;  Mismatches  117;  Indels   97;  Gaps   15;

Qy        125 FGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSG-YPGCPYP- 182
              ||: ||  ::|:|    |||    ||  :  :|:||        ||| |    || | |
Db         89 FGQPPP--TQPSV----PPYGVYPPPGGNPPSGVPS--------YPPFPGAPVPGQPMPP 134

Qy        183 ----PAGPYPATTSSQYPSQPPVTTVG----PSRDGTISEDTI----------RASLIS 223
                  | ||||      || | ||   |    ||  |     |:           | ::
Db        135 PGHQPPGPYPGQLPVTYPGQSPVPPPGQQPMPSYPGYPGSGTVTPAVPPVQFGNRGTITD 194

Qy        224 AV-------SDKLRWRMKEEMDGAQAELNAL-----KRTEEDLKK-----GHQKLEEMVT 266
              |        :: ||  ||      || :: |     |: :: |       |   :::: :
Db        195 ASGFDPLRDAEVLRKAMKGFGTDEQAIIDCLGSRSNKQRQQILLSFKTAYGKDLIKDLKS 254

Qy        267 RLD------------------QEVAEVDKN--------IELLKKKDEELSSALEKMENQ 299
              |                    |: |  |         ||:|  :  |    | |
Db        255 ELSGNFEKTILALMKTPILFDAYEIKEAIKGAGTDEACLIEILASRSNEHIRELNKAYKT 314

Qy        300 SENNDIDEVI-IPTAPLYKQILNLYAEENAIEDT------------IFYLGEALRRGVI 345
                  ::| |   |:  ::::|   :: |  | |             ::  ||      :
Db        315 EFKKTLEEAIRSDTSGHFQRLLISLSQGNRDESTNVDMSLVQRDVQELYAAGE----NRL 370

Qy        346 DLDVFLKHVRLLSRKQFQLRALMQKARKTAG 376
                |    :  | || :  | |:  : ::  |
Db        371 GTDESKFNAVLCSRSRAHLVAVFNEYQRMTG 401


RESULT 7
ANX7_MOUSE
ID   ANX7_MOUSE     STANDARD;      PRT;   463 AA.
AC   Q07076;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Annexin A7 (Annexin VII) (Synexin).
GN   ANXA7 OR ANX7.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93168121; PubMed=7916616;
RA   Zhang-Keck Z.Y., Burns A.L., Pollard H.B.;
RT   "Mouse synexin (annexin VII) polymorphisms and a phylogenetic
RT   comparison with other synexins.";
RL   Biochem. J. 289:735-741(1993).
CC   -!- FUNCTION: CALCIUM/PHOSPHOLIPID-BINDING PROTEIN WHICH PROMOTES
CC       MEMBRANE FUSION AND IS INVOLVED IN EXOCYTOSIS.
CC   -!- DOMAIN: A pair of annexin repeats may form one binding site for
CC       calcium and phospholipid.
CC   -!- SIMILARITY: BELONGS TO THE ANNEXIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 4 ANNEXIN REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L13129; AAA37238.1; -.
DR   HSSP; P26256; 1DM5.
DR   MGD; MGI:88031; Anxa7.
DR   InterPro; IPR001464; Annexin.
DR   Pfam; PF00191; annexin; 4.
DR   PRINTS; PR00196; ANNEXIN.
DR   ProDom; PD000143; Annexin; 4.
DR   SMART; SM00335; ANX; 4.
DR   PROSITE; PS00223; ANNEXIN; 4.
KW   Annexin; Calcium/phospholipid-binding; Repeat.
FT   DOMAIN        1    143       REPEAT-RICH REGION.
FT   REPEAT      169    229       ANNEXIN 1.
FT   REPEAT      241    301       ANNEXIN 2.
FT   REPEAT      324    384       ANNEXIN 3.
FT   REPEAT      400    460       ANNEXIN 4.
FT   DOMAIN        5     20       3 X 5 AA TANDEM REPEATS OF G-Y-P-P-X.
FT   REPEAT        5      9       1.
FT   REPEAT       10     14       2.
FT   REPEAT       16     20       3.
SQ   SEQUENCE   463 AA;  49939 MW;  50F7B20FD48EBCD5 CRC64;

  Query Match             6.8%;  Score 135.5;  DB 1;  Length 463;
  Best Local Similarity   21.5%;  Pred. No. 0.57;
  Matches   74;  Conservative   42;  Mismatches  123;  Indels  105;  Gaps   15;

Qy        126 GEEP-PVFSRPTVSASYPPYTATGPPNTSYMPG-------------MPSGISAYPSGY- 169
              |: | |    |    :|||  : | |     |                 | |  ||| |
Db         30 GQYPYPSGFPPMGGGAYPPAPSGGYPGAGGYPAPGGYPAPGGYPGALSPGGPPAYPGGQG 89

Qy        170 ---PPNPSGYPGCPYPPA-----GP-----------------YPATTSSQYPSQP----- 199
                ||  :|: | | |||      ||                 ||    : |||||
Db         90 FGAPPGGAGFSGYPQPPAQSYGGGPAQVPVPGGFPGGQMPSQYPG-GQAPYPSQPASMTQ 148

Qy        200 -------PVTTVGPSRDGTISEDTIRA------SLISAVSDK----------------- 228
                    | :     ||  |    ::       :::  ||::
Db        149 GTQGTILPASNFDAMRDAEILRKAMKGFGTDEQAIVDVVSNRSNDQRQQIKAAFKTMYGK 208

Qy        229 -LRWRMKEEMDGAQAELNA---LKRTEED---LKKGHQ--------KLEEMVTRLDQEVA 273
               |   :| |: | ||     :  | |   |:|  |         :| : || :||:
Db        209 DLIKDLKSELSGNMEELILALFMPSTYYDAWSLRKAMQGAGTQERVLIEILCTRTNQEIR 268
```

```
Qy        274 EVDKNIELLKKKDEELSSALEKMENQSENNDIDEVIIPTAPLYKQILNLYAEENAIEDT- 332
              :: :  :|      |    |||      :   : :::      :          : | ||
Db        269 DIVRCYQL------EFGRDLEKDIRSDTSGHFERLLVSMCQANRDERQSVNHQMAQEDAQ 322

Qy        333 -IFYLGEALRRGVIDLDVFLKHVRLLSRKQFQLRALMQKARKTA 375
               ::  ||    | :  |    :: | :|   ||:| |:   : |
Db        323 RLYQAGE----GRLGTDESCFNMILATRSFPQLKATMEAYSRMA 362


RESULT 8
WASL_BOVIN
ID   WASL_BOVIN     STANDARD;      PRT;   505 AA.
AC   Q95107;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Neural Wiskott-Aldrich syndrome protein (N-WASP).
GN   WASL.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=97050838; PubMed=8895577;
RA   Miki H., Miura K., Takenawa T.;
RT   "N-WASP, a novel actin-depolymerizing protein, regulates the cortical
RT   cytoskeletal rearrangement in a PIP2-dependent manner downstream of
RT   tyrosine kinases.";
RL   EMBO J. 15:5326-5335(1996).
CC   -!- FUNCTION: REGULATES ACTIN POLYMERIZATION BY STIMULATING THE ACTIN-
CC       NUCLEATING ACTIVITY OF THE ACTIN-RELATED PROTEIN 2/3 (ARP2/3)
CC       COMPLEX.
CC   -!- SUBUNIT: BINDS ACTIN AND ARP2/3 COMPLEX; INTERACTS WITH CDC42
CC       BINDS TO SH3 DOMAINS OF ASH/GRB2.
CC   -!- SIMILARITY: CONTAINS 1 CRIB DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 WH1 DOMAIN.
CC   -!- SIMILARITY: CONTAINS 2 WH2 DOMAINS.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; D67066; BAA11082.1; -.
DR   InterPro; IPR000095; PAKbox/Rhobndng.
DR   InterPro; IPR000697; RanBP1_WASP.
DR   InterPro; IPR001960; WH1.
DR   InterPro; IPR003124; WH2.
DR   Pfam; PF00568; WH1; 1.
DR   Pfam; PF00786; PBD; 1.
DR   Pfam; PF02205; WH2; 2.
DR   SMART; SM00285; PBD; 1.
DR   SMART; SM00461; WH1; 1.
DR   SMART; SM00246; WH2; 2.
DR   PROSITE; PS50108; CRIB; 1.
KW   Actin-binding; Repeat.
FT   DOMAIN       34     138       WH1.
FT   DOMAIN      203     216       CRIB.
FT   DOMAIN      277     392       PRO-RICH.
FT   DOMAIN      405     422       WH2 1.
FT   DOMAIN      433     450       WH2 2.
FT   DOMAIN      486     505       ASP-RICH.
SQ   SEQUENCE   505 AA;  54671 MW;  54B83B48F1CDB3B8 CRC64;

  Query Match             6.7%;  Score 133;  DB 1;  Length 505;
  Best Local Similarity   27.0%;  Pred. No. 0.87;
  Matches   58;  Conservative   31;  Mismatches   66;  Indels   60;  Gaps   12;

Qy        126 GEEPPVFSR-PTVSASYPPYTATG------PPNTSYMPGMPSGISAYPSGYPPNP----- 173
              |  ||  || || :   || :  |       |||  | | :|:  |: ||| || |
Db        308 GAPPPPPSRAPTAAPPPPPPSRPGVGAPPPPPNRMYPPPLPALPSSAPSGPPPPPPPLSV 367

Qy        174 SGYPGCPYPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRM 233
              ||    | ||  | |       |  ||    |   ||      |    |  :|: |::|
Db        368 SGSVAPPPPPPPPP-------PGPPP--PPGLPSDGDHQVPTPAGSK-AALLDQIR--- 414

Qy        234 KEEMDGAQAELNALKRTEEDLK----KGHQKLEEMVTRLDQ--EVAEVDKN--------- 278
                  :|||     ||: |:: :     |   | : : :  |   | :  ::
Db        415 ----EGAQ-----LKKVEQNSRPVSCSGRDALLDQIRQGIQLKSVTDAPESTPPAPAPTS 465

Qy        279 ------IELLKKKDEELSSALEKMENQSENNDIDE 307
                   :|:::|: : : |:     :: |: | ||
Db        466 GIVGALMEVMQKRSKAIHSS-----DEDEDEDDDE 495


RESULT 9
DYNA_RAT
ID   DYNA_RAT       STANDARD;      PRT;  1280 AA.
AC   P28023;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Dynactin 1 (150 kDa dynein-associated polypeptide) (DP-150) (DAP-150)
DE   (p150-glued).
GN   DCTN1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley;
RX   MEDLINE=91260877; PubMed=1828535;
RA   Holzbaur E.L.F., Hammarback J.A., Paschal B.M., Kravit N.G.,
RA   Pfister K.K., Vallee R.B.;
RT   "Homology of a 150K cytoplasmic dynein-associated polypeptide with
RT   the Drosophila gene Glued.";
RL   Nature 351:579-580(1991).
RN   [2]
RP   REVISIONS.
RA   Holzbaur E.L.F., Hammarback J.A., Paschal B.M., Kravit N.G.,
RA   Pfister K.K., Vallee R.B.;
RL   Submitted (DEC-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: REQUIRED FOR THE CYTOPLASMIC DYNEIN-DRIVEN RETROGRADE
CC       MOVEMENT OF VESICLES AND ORGANELLES ALONG MICROTUBULES. DYNEIN-
CC       DYNACTIN INTERACTION IS A KEY COMPONENT OF THE MECHANISM OF AXONAL
CC       TRANSPORT OF VESICLES AND ORGANELLES.
CC   -!- SUBUNIT: LARGE MACROMOLECULAR COMPLEX OF AT LEAST 10 COMPONENTS.
CC       P150(GLUED) BINDS DIRECTLY TO MICROTUBULES AND TO CYTOPLASMIC
CC       DYNEIN.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: BELONGS TO THE DYNACTIN 150 KDA SUBUNIT FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 CAP-GLY DOMAIN.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X62160; CAA44091.1; -.
DR   PIR; S16129; S16129.
DR   InterPro; IPR000938; CAP-Gly.
DR   Pfam; PF01302; CAP_GLY; 1.
DR   PROSITE; PS00845; CAP_GLY_1; 1.
DR   PROSITE; PS50245; CAP_GLY_2; 1.
KW   Motor protein; Microtubules; Dynein; Coiled coil; Cytoskeleton.
FT   DOMAIN       48     90       CAP-GLY.
```

```
FT   DOMAIN       157    184       SER-RICH.
FT   DOMAIN       214    513       COILED COIL (POTENTIAL).
FT   DOMAIN       942   1048       COILED COIL (POTENTIAL).
FT   DOMAIN      1184   1213       COILED COIL (POTENTIAL).
SQ   SEQUENCE   1280 AA;  141929 MW;  C9348CF129F4FF5C CRC64;

  Query Match             6.6%;  Score 131;  DB 1;  Length 1280;
  Best Local Similarity   25.4%;  Pred. No. 3.4;
  Matches   72;  Conservative   45;  Mismatches  100;  Indels   66;  Gaps   16;

Qy        134 RPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCPYPPAGPYPATTSS 193
              :||  ||    |    |::|  |   :      |  |  |: |    |  | || ||
Db        151 KPTRPAS----TGVAGPSSSLGPSGSASAGELSSSEPSTPAQTPLA--APIIPTPALTSP 204

Qy        194 QYPSQPPVTTVGPSRDG------------TI----SEDTIRASLISAVSDKLR------ 230
              : ||: :     :|             |:    |||  :| |      |::
Db        205 --GAAPPLPSPSKEEEGLRDQVRDLEEKLETLRLKRSED--KAKLKELEKHKIQLEQVQE 260

Qy        231 W--RMKEEMDGAQAELNALKRTEEDLKKGHQKLEEM--------VTRLDQEVAE-----V 275
              |  :|:|:    |  |   |  :| |:   : :|||        :  ||:|:||     :
Db        261 WKSKMQEQQADLQRRLKEAKEAKEALEAKERYMEEMADTADAIEMATLDKEMAEERAESL 320

Qy        276 DKNIELLKKKDEELSSALEKMENQSENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFY 335
              : :| ||:: :||:: || :: : |    |      |  |:  |    |:||
Db        321 QQEVEALKERVDELTTDLEILKAEIEEKGSD----GAASSYQ--LKQLEEQNA------R 368

Qy        336 LGEALRRGVIDLDVFLK--HV---RLLSRKQFQLRALMQKARK 373
              | :|| | : ||    |  ||   :|: :|  :|  : |:  :
Db        369 LKDALVR-MRDLSSSEKQEHVKLQKLMEKKNQELEVVRQQRER 410
```

```
RESULT 10
ANXB_HUMAN
ID   ANXB_HUMAN     STANDARD;      PRT;   505 AA.
AC   P50995;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Annexin A11 (Annexin XI) (Calcyclin-associated annexin 50) (CAP-50)
DE   (56 kDa autoantigen).
GN   ANXA11 OR ANX11.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Teratocarcinoma;
RX   MEDLINE=94140847; PubMed=7508441;
RA   Misaki Y., Pruijn G.J., van der Kemp A.W., van Venrooij W.J.;
RT   "The 56K autoantigen is identical to human annexin XI.";
RL   J. Biol. Chem. 269:4240-4246(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20469408; PubMed=11013079;
RA   Bances P., Fernandez M.R., Rodriguez-Garcia M.I., Morgan R.O.,
RA   Fernandez M.P.;
RT   "Annexin A11 (ANXA11) gene structure as the progenitor of paralogous
RT   annexins and source of orthologous cDNA isoforms.";
RL   Genomics 69:95-103(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lymph;
RA   Strausberg R.;
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC AND POSSIBLY NUCLEAR.
CC   -!- DOMAIN: A pair of annexin repeats may form one binding site for
CC       calcium and phospholipid.
CC   -!- DISEASE: ANTIBODIES AGAINST ANXA11 ARE PRESENT IN SERA FROM
CC       PATIENTS WITH VARIOUS AUTOIMMUNE DISEASES, PREDOMINANTLY IN SERA
CC       FROM PATIENTS WITH RHEUMATOID ARTHRITIS, SYSTEMIC LUPUS
CC       ERYTHEMATOSUS, OR SJOGREN'S SYNDROME.
CC   -!- SIMILARITY: BELONGS TO THE ANNEXIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 4 ANNEXIN REPEATS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L19605; AAA19734.1; -.
DR   EMBL; AJ278463; CAB94995.1; -.
DR   EMBL; AJ278464; CAB94996.1; -.
DR   EMBL; AJ278465; CAB94997.1; -.
DR   EMBL; BC007564; AAH07564.1; -.
DR   HSSP; P13214; 1AOW.
DR   Genew; HGNC:535; ANXA11.
DR   MIM; 602572; -.
DR   InterPro; IPR001464; Annexin.
DR   Pfam; PF00191; annexin; 4.
DR   PRINTS; PR00196; ANNEXIN.
DR   ProDom; PD000143; Annexin; 4.
DR   SMART; SM00335; ANX; 4.
DR   PROSITE; PS00223; ANNEXIN; 4.
KW   Annexin; Calcium/phospholipid-binding; Repeat; Polymorphism.
FT   REPEAT       209    269       ANNEXIN 1.
FT   REPEAT       281    341       ANNEXIN 2.
FT   REPEAT       365    425       ANNEXIN 3.
FT   REPEAT       440    500       ANNEXIN 4.
FT   VARIANT      230    230       R -> C (IN DBSNP:1049550).
FT                                 /FTId=VAR_012006.
FT   VARIANT      457    457       I -> V (IN DBSNP:1802932).
FT                                 /FTId=VAR_012007.
SQ   SEQUENCE   505 AA;  54389 MW;  4ADCAC8F270BFEE4 CRC64;

  Query Match             6.4%;  Score 128;  DB 1;  Length 505;
  Best Local Similarity   28.1%;  Pred. No. 1.7;
  Matches   55;  Conservative   15;  Mismatches   76;  Indels   50;  Gaps    9;

Qy        139 ASYPPYT--ATGPPNTSYMPGMPSGISAYPSGYPPN--PS--GYPGCPYP---------- 182
              |.|||      | | ::  |  | |:   | | ||:  ||   ||| | |
Db         80 AGYPPVPPGGFGQPPSAQQPVPPYGMYPPPGGNPPSRMPSYPPYPGAPVPGQPMPPPGQQ 139

Qy        183 PAGPYPATTSSQYPSQPPVTTVG-----PSRDGTISEDTI----------RASLISAV- 225
              | | ||      || ||||   |     ||  |    |:          | ::  |
Db        140 PPGAYPGQPPVTYPGQPPVPLPGQQQPVPSYPGYPGSGTVTPAVPPTQFGSRGTITDAPG 199

Qy        226 ------SDKLRWRMKEEMDGAQAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNI 279
                   :: ||  ||     || :: |       |:  | |    |  :
Db        200 FDPLRDAEVLRKAMKGFGTDEQAIIDCLGSRSN--KQRQQILLSFKTAYGK--------- 248

Qy        280 ELLKKKDEELSSALEK 295
              :|:|    |||   ||
Db        249 DLIKDLKSELSGNFEK 264


RESULT 11
DYNA_MOUSE
ID   DYNA_MOUSE     STANDARD;      PRT;  1281 AA.
AC   O08788;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Dynactin 1 (150 kDa dynein-associated polypeptide) (DP-150) (DAP-150)
DE   (p150-glued).
GN   DCTN1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J;
```

```
RX   MEDLINE=97223454; PubMed=9070275;
RA   Jang W., Weber J.S., Tokito M.K., Holzbaur E.L., Meisler M.H.;
RT   "Mouse p150Glued (dynactin 1) cDNA sequence and evaluation as a
RT   candidate for the neuromuscular disease mutation mnd2.";
RL   Biochem. Biophys. Res. Commun. 231:344-347(1997).
CC   -!- FUNCTION: REQUIRED FOR THE CYTOPLASMIC DYNEIN-DRIVEN RETROGRADE
CC       MOVEMENT OF VESICLES AND ORGANELLES ALONG MICROTUBULES. DYNEIN-
CC       DYNACTIN INTERACTION IS A KEY COMPONENT OF THE MECHANISM OF AXONAL
CC       TRANSPORT OF VESICLES AND ORGANELLES.
CC   -!- SUBUNIT: LARGE MACROMOLECULAR COMPLEX OF AT LEAST 10 COMPONENTS.
CC       P150(GLUED) BINDS DIRECTLY TO MICROTUBULES AND TO CYTOPLASMIC
CC       DYNEIN.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: BELONGS TO THE DYNACTIN 150 KDA SUBUNIT FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 CAP-GLY DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U60312; AAB57773.1; -.
DR   MGD; MGI:107745; Dctn1.
DR   InterPro; IPR000938; CAP-Gly.
DR   Pfam; PF01302; CAP_GLY; 1.
DR   PROSITE; PS00845; CAP_GLY_1; 1.
DR   PROSITE; PS50245; CAP_GLY_2; 1.
KW   Motor protein; Microtubules; Dynein; Coiled coil; Cytoskeleton.
FT   DOMAIN       48     90       CAP-GLY.
FT   DOMAIN      157    184       SER-RICH.
FT   DOMAIN      214    547       COILED COIL (POTENTIAL).
FT   DOMAIN      943   1049       COILED COIL (POTENTIAL).
FT   DOMAIN     1185   1214       COILED COIL (POTENTIAL).
SQ   SEQUENCE   1281 AA;  141721 MW;  3087FBFF0847D1EC CRC64;

  Query Match             6.4%;  Score 127.5;  DB 1;  Length 1281;
  Best Local Similarity   24.5%;  Pred. No. 5.4;
  Matches   69;  Conservative   46;  Mismatches  104;  Indels   63;  Gaps   14;

Qy        134 RPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCPYPPAGPYPATTSS 193
              :||  ||    |    |::|  |   :      |  |  |:  |     |  | || ||
Db        151 KPTRPAS----TGVAGPSSSLGPSGSASAGELSSSEPSTPAQTPLA--APIIPTPALTSP 204

Qy        194 QYPSQPPVTTVGPSRDGTIS-----EDTIRASLISAVSDKLRWR-------------MK 234
              : ||: :     :|  :     |: :    :    || : :                |
Db        205 --GAAPPLPSPSKEEEGLRAQVRDLEEKLETLRLKRSEDKAKLKELEKHKIQLEQVQEWK 262

Qy        235 EEMDGAQAEL-----NALKRTEEDLKKGHQKLEEM--------VTRLDQEVAE-----VD 276
              :|   ||:|      | |  :| |:   : :|||        :  ||:|:||     :
Db        263 SKMQEQQADLQRRLKEARKEAKEALEAKERYMEEMADTADAIEMATLDKEMAEERAESLQ 322

Qy        277 KNIELLKKKDEELSSALEKMENQSENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYL 336
              : :| ||:: :||:: || :: : |    |      |  |:  |    |:||       |
Db        323 QEVEALKERVDELTTDLEILKAEIEEKGSD----GAASSYQ--LKQLEEQNA------RL 370

Qy        337 GEALRRGVIDLDVFLK--HV---RLLSRKQFQLRALMQKARK 373
              :|| | : ||    |  ||   :|: :|  :|  : |:  :
Db        371 KDALVR-MRDLSSSEKQEHVKLQKLMEKKNQELEVVRQQRER 411


RESULT 12
DYNA_CHICK
ID   DYNA_CHICK     STANDARD;      PRT;  1224 AA.
AC   P35458;
DT   01-JUN-1994 (Rel. 29, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Dynactin 1 (150 kDa dynein-associated polypeptide) (DP-150) (DAP-150)
DE   (p150-glued).
GN   DCTN1.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Cleveland D.W.;
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 172-1224 FROM N.A.
RC   TISSUE=Embryonic brain;
RX   MEDLINE=92098576; PubMed=1836789;
RA   Gill S.R., Schroer T.A., Szilak I., Steuer E.R., Sheetz M.P.,
RA   Cleveland D.W.;
RT   "Dynactin, a conserved, ubiquitously expressed component of an
RT   activator of vesicle motility mediated by cytoplasmic dynein.";
RL   J. Cell Biol. 115:1639-1650(1991).
CC   -!- FUNCTION: DYNACTIN IS A MAJOR COMPONENT OF ACTIVATOR I, A 20S
CC       POLYPEPTIDE COMPLEX THAT STIMULATES DYNEIN-MEDIATED VESICLE
CC       TRANSPORT.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS ARE PRODUCED BY
CC       ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED.
CC   -!- SIMILARITY: BELONGS TO THE DYNACTIN 150 KDA SUBUNIT FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 CAP-GLY DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X62773; CAA44617.2; -.
DR   PIR; A41642; A41642.
DR   InterPro; IPR000938; CAP-Gly.
DR   Pfam; PF01302; CAP_GLY; 1.
DR   PROSITE; PS00845; CAP_GLY_1; 1.
DR   PROSITE; PS50245; CAP_GLY_2; 1.
KW   Motor protein; Microtubules; Dynein; Coiled coil; Cytoskeleton;
KW   Alternative splicing.
FT   DOMAIN       49     91       CAP-GLY.
FT   DOMAIN      205    540       COILED COIL (POTENTIAL).
FT   DOMAIN      936   1042       COILED COIL (POTENTIAL).
FT   DOMAIN     1081   1117       COILED COIL (POTENTIAL).
SQ   SEQUENCE   1224 AA;  135562 MW;  03B7FFE68E7C01D7 CRC64;

  Query Match             6.4%;  Score 127;  DB 1;  Length 1224;
  Best Local Similarity   26.3%;  Pred. No. 5.5;
  Matches   76;  Conservative   39;  Mismatches   90;  Indels   84;  Gaps   16;

Qy        145 TATGPPNTSYMPGMPSGISAYPSG-----------YPPNPSGYPGCPYPPAGPYPATTS 192
              ||  |  |      ||  :| |||            |  |:  |     |  | |: ||
Db        140 TARRPKPTRTPTSAPSSGTAGPSGSASASGGEMSSSEPSTPAQTP--LVAPVIPSPSLTS 197

Qy        193 SQYPSQPPVTTVGPSRDGTISEDTIRASL----------ISAVSDKLRWR--------- 232
                  ||  : ||   |  |: :|: :           |    || : :
Db        198 -------PVAPMVPS--PTKEEENLRSQVRDLEEKLETLKIKRNEDKAKLKELEKYKIQL 248

Qy        233 -----MKEEMDGAQAEL-----NALKRTEEDLKKGHQKLEEM--------VTRLDQEVAE 274
                  | :|   ||:|      | |  :: |:   : :|||         :  ||:|:||
Db        249 EQVQEWKSKMQEQQADLQRRLKEAKKEAKDALEAKERYMEEMADTADAIEMATLDKEMAE 308

Qy        275 -----VDKNIELLKKKDEELSSALEKMENQSENNDIDEVIIPTAPLYKQILNLYAEENAI 329
                  : : :: ||:| | |:  || :::: |    |      |  | |:  |  |:||
Db        309 ERAESLQQEVDSLKEKVEYLTMDLEILKHEIEEKGSD----GAASSY-QVKQL-EEQNA- 361

Qy        330 EDTIFYLGEALRRGVIDLDVFLK--HVRL---LSRKQFQLRALMQKARK 373
                  | ||| | : ||    |  ||:|   : :|  :| :| |:  |
Db        362 -----RLKEALVR-MRDLSASEKQEHVKLQKQMEKKNTELESLRQQREK 404
```

```
RESULT 13
WASL_HUMAN
ID   WASL_HUMAN     STANDARD;      PRT;   505 AA.
AC   O00401;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Neural Wiskott-Aldrich syndrome protein (N-WASP).
GN   WASL.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=97464048; PubMed=9322739;
RA   Fukuoka M., Miki H., Takenawa T.;
RT   "Identification of N-WASP homologs in human and rat brain.";
RL   Gene 196:43-48(1997).
CC   -!- FUNCTION: REGULATES ACTIN POLYMERIZATION BY STIMULATING THE ACTIN-
CC       NUCLEATING ACTIVITY OF THE ACTIN-RELATED PROTEIN 2/3 (ARP2/3)
CC       COMPLEX.
CC   -!- SUBUNIT: BINDS ACTIN AND ARP2/3 COMPLEX; INTERACTS WITH CDC42
CC       BINDS TO SH3 DOMAINS OF ASH/GRB2.
CC   -!- SIMILARITY: CONTAINS 1 CRIB DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 WH1 DOMAIN.
CC   -!- SIMILARITY: CONTAINS 2 WH2 DOMAINS.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; D88460; BAA20128.1; -.
DR   Genew; HGNC:12735; WASL.
DR   MIM; 605056; -.
DR   InterPro; IPR000095; PAKbox/Rhobndng.
DR   InterPro; IPR000697; RanBP1_WASP.
DR   InterPro; IPR001960; WH1.
DR   InterPro; IPR003124; WH2.
DR   Pfam; PF00568; WH1; 1.
DR   Pfam; PF00786; PBD; 1.
DR   Pfam; PF02205; WH2; 2.
DR   SMART; SM00285; PBD; 1.
DR   SMART; SM00461; WH1; 1.
DR   SMART; SM00246; WH2; 2.
DR   PROSITE; PS50108; CRIB; 1.
KW   Actin-binding; Repeat.
FT   DOMAIN       34    138       WH1.
FT   DOMAIN      203    216       CRIB.
FT   DOMAIN      277    392       PRO-RICH.
FT   DOMAIN      405    422       WH2 1.
FT   DOMAIN      433    450       WH2 2.
FT   DOMAIN      485    505       ASP-RICH.
SQ   SEQUENCE   505 AA;  54824 MW;  BC5670A11AB63539 CRC64;

  Query Match             6.3%;  Score 126.5;  DB 1;  Length 505;
  Best Local Similarity   27.5%;  Pred. No. 2;
  Matches   56;  Conservative   31;  Mismatches   64;  Indels   53;  Gaps   12;

Qy        129 PPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGY----PGCPYPPA 184
              ||  |||:|    ||     |||  | |  |:  |: ||| || |       |  | ||
Db        324 PPPPSRPSVEVP-PP-----PPNRMYPPPPPALPSSAPSGPPPPPPSVLGVGPVAPPPPP 377

Qy        185 GPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGAQAEL 244
              | |       |  ||     || ||     |  |   :|: |::|       :|||
Db        378 PPPP-------PPGPPPPPGLPS-DGDHQVPT-TAGNKAALLDQIR-------EGAQ--- 418

Qy        245 NALKRTEEDLK----KGHQKLEEMVTRLDQ--EVAE--------------VDKNIELLK 283
                ||: |:: :     |   | : : :  |   ||:            |   :|:::
Db        419 --LKKVEQNSRPVSCSGRDALLDQIRQGIQLKSVADGQESTPPTPAPTSGIVGALMEVMQ 476

Qy        284 KKDEELSSALEKMENQSENNDIDE 307
              |: : : |: :: |:: :  | ::
Db        477 KRSKAIHSS-DEDEDEDDEEDFED 499


RESULT 14
VASP_MOUSE
ID   VASP_MOUSE     STANDARD;      PRT;   376 AA.
AC   P70460; Q9R214;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Vasodilator-stimulated phosphoprotein (VASP).
GN   VASP.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129;
RX   MEDLINE=96411679; PubMed=8812448;
RA   Zimmer M., Fink T., Fischer L., Hauser W., Scherer K., Lichter P.,
RA   Walter U.;
RT   "Cloning of the VASP (vasodilator-stimulated phosphoprotein) genes in
RT   human and mouse: structure, sequence, and chromosomal localization.";
RL   Genomics 36:227-233(1996).
RN   [2]
RP   SEQUENCE FROM N.A., AND PHOSPHORYLATION.
RC   STRAIN=C57BL/6J;
RX   MEDLINE=99185054; PubMed=10085070;
RA   Collins S.P., Uhler M.D.;
RT   "Cyclic AMP- and cyclic GMP-dependent protein kinases differ in their
RT   regulation of cyclic AMP response element-dependent gene
RT   transcription.";
RL   J. Biol. Chem. 274:8391-8404(1999).
CC   -!- FUNCTION: ACTIN- AND PROFILIN-BINDING MICROFILAMENT-ASSOCIATED
CC       PROTEIN. MAY ACT IN CONCERT WITH PROFILIN TO CONVEY SIGNAL
CC       TRANSDUCTION TO ACTIN FILAMENT PRODUCTION (BY SIMILARITY).
CC   -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: FOCAL ADHESIONS (BY SIMILARITY).
CC   -!- PTM: MAJOR SUBSTRATE FOR CAMP-DEPENDENT (CAPK) AND CGMP-DEPENDENT
CC       PROTEIN KINASE (CGPK) IN PLATELETS.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X98475; CAA67108.1; -.
DR   EMBL; AF084548; AAD16045.1; -.
DR   MGD; MGI:109268; Vasp.
DR   InterPro; IPR000697; RanBP1_WASP.
DR   InterPro; IPR001960; WH1.
DR   Pfam; PF00568; WH1; 1.
DR   SMART; SM00461; WH1; 1.
KW   Phosphorylation; Actin-binding.
FT   DOMAIN       166    182       POLY-PRO.
FT   DOMAIN       318    321       POLY-SER.
FT   MOD_RES      153    153       PHOSPHORYLATION (BY PKA AND PKG) (BY
FT                                 SIMILARITY).
FT   MOD_RES      235    235       PHOSPHORYLATION (BY PKA AND PKG) (BY
FT                                 SIMILARITY).
FT   MOD_RES      274    274       PHOSPHORYLATION (BY PKA AND PKG) (BY
FT                                 SIMILARITY).
```

```
FT   CONFLICT    209    209       T -> A (IN REF. 2).
FT   CONFLICT    288    288       MISSING (IN REF. 2).
SQ   SEQUENCE   376 AA;  39824 MW;  CC338D07519A0294 CRC64;

  Query Match             6.2%;  Score 123.5;  DB 1;  Length 376;
  Best Local Similarity   20.4%;  Pred. No. 2.1;
  Matches   80;  Conservative   56;  Mismatches  121;  Indels  135;  Gaps   18;

Qy         33 VFNDGSSRELVNLTGTIPVRY-RGNIYNIPICLWLLDTYPYNPPICFVKPTSSMTIKTGK 91
              :::| : | |    || |  : |  ||:                      ||::     |:
Db         15 LYDDSNKRWLP--AGTGPQAFSRVQIYH--------------------NPTANSFRVVGR 52

Qy         92 HVDANGKIYL-------------PYLHDWKHPRSELLELIQIMIVIFGEEPPVFSRPTV 137
               :  : :: :             |  | |:  |       |:  : || :       |
Db         53 KMQPDQQVVINCAIIRGVKYNQATPIFHQWRDAR-------QVWGLNFGSKEDAIQFATG 105

Qy        138 SASY--------PPYTATGPPNTSYMPGMPS----------------------GISAYPS 167
              |:         ||  |  ||  |   | ||                      |  | |:
Db        106 MANALEALEGGGPP-PAPAPPAWSAQNG-PSPEELEQQKRQPEHMERRVSNAGGPPAPPA 163

Qy        168 GYPPNPSGYPGCPYPPAGPYPATTSSQY-----------PSQPPV-TTVGPSRDGTISED 215
              | || |·| |  | ||  | |   ||             |  ||: || ||:  |:
Db        164 GGPPPPPGPPPPPGPP--PPPGLPSSGVSGAGHGAGAAPPPAPPLPTTQGPNSGGS-GAP 220

Qy        216 TIRASLISAVSDKLRWRMKEEM----------------DGAQAELNA-LKRTEEDLKKG 257
              : |::  |   |||   |:|                 |   |:|| | |  :  : |
Db        221 GLAAAIAGA---KLRKVSKQEEASGGPLAPKAENSRSTGGGLMEEMNAMLARRRKATQVG 277

Qy        258 HQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMEN--------------QSEN 302
              :  ::     ::  | :   | :::  |: |: | :|::                |::
Db        278 EKPPKDESASQEESEARLPAQSEPVRRPWEKNSTTLPRMKSSSSVTTSEAHPSTPCSSDD 337

Qy        303 ND--------IDEVIIPTAPLYKQILNLYAEE 326
              :|        ::||      : ::|: :: :|
Db        338 SDLERVKQELLEEVRKELQKMKEEIIEVFVQE 369


RESULT 15
WASL_RAT
ID   WASL_RAT       STANDARD;      PRT;   501 AA.
AC   O08816;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Neural Wiskott-Aldrich syndrome protein (N-WASP).
GN   WASL.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97464048; PubMed=9322739;
RA   Fukuoka M., Miki H., Takenawa T.;
RT   "Identification of N-WASP homologs in human and rat brain.";
RL   Gene 196:43-48(1997).
CC   -!- FUNCTION: REGULATES ACTIN POLYMERIZATION BY STIMULATING THE ACTIN-
CC       NUCLEATING ACTIVITY OF THE ACTIN-RELATED PROTEIN 2/3 (ARP2/3)
CC       COMPLEX (BY SIMILARITY).
CC   -!- SUBUNIT: BINDS ACTIN AND ARP2/3 COMPLEX; INTERACTS WITH CDC42
CC       BINDS TO SH3 DOMAINS OF ASH/GRB2 (BY SIMILARITY).
CC   -!- SIMILARITY: CONTAINS 1 CRIB DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 WH1 DOMAIN.
CC   -!- SIMILARITY: CONTAINS 2 WH2 DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D88461; BAA21534.1; -.
DR   InterPro; IPR000095; PAKbox/Rhobndng.
DR   InterPro; IPR000697; RanBP1_WASP.
DR   InterPro; IPR001960; WH1.
DR   InterPro; IPR003124; WH2.
DR   Pfam; PF00568; WH1; 1.
DR   Pfam; PF00786; PBD; 1.
DR   Pfam; PF02205; WH2; 2.
DR   SMART; SM00285; PBD; 1.
DR   SMART; SM00461; WH1; 1.
DR   SMART; SM00246; WH2; 2.
DR   PROSITE; PS50108; CRIB; 1.
KW   Actin-binding; Repeat.
FT   DOMAIN      31    135       WH1.
FT   DOMAIN     200    213       CRIB.
FT   DOMAIN     274    385       PRO-RICH.
FT   DOMAIN     401    418       WH2 1.
FT   DOMAIN     429    446       WH2 2.
FT   DOMAIN     482    501       ASP-RICH.
SQ   SEQUENCE   501 AA;  54325 MW;  480E21F26F7FC77E CRC64;

  Query Match             6.1%;  Score 122.5;  DB 1;  Length 501;
  Best Local Similarity   27.9%;  Pred. No. 3.4;
  Matches   60;  Conservative   28;  Mismatches   60;  Indels   67;  Gaps   14;

Qy        126 GEEPPVFSR-PTVSASYPPYTATG------PPNTSYMPGMPSGISAYPSGYPPNP----S 174
              |  ||  || || :   || :  |      |||  | |  |:  |: ||| || |    :
Db        305 GAPPPPPSRAPTAAPPPPPPSRPGVVVPPPPPNRMYPPPPPALPSSAPSGPPPPPPLSMA 364

Qy        175 GYPGCPYPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASL--ISAVSDKLRWR 232
              |    | ||  | |       |  ||    |  ||   :  : ||    :|: |::|
Db        365 GSTAPPPPPPPPP-------PGPPP--PPGLPSDG---DHQVPASSGNKAALLDQIR-- 410

Qy        233 MKEEMDGAQAELNALKRTEEDLK----KGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEE 288
                  :|||     ||: |:: :     |   |      |||    : : |:|    | :
Db        411 -----EGAQ-----LKKVEQNSRPVSCSGRDAL------LDQ----IRQGIQLKSVSDGQ 450

Qy        289 -------------LSSALEKMENQSE---NNDIDE 307
                          : : :| |: :|:   ::| ||
Db        451 ESTPPTPAPTSGIVGALMEVMQKRSKAIHSSDEDE 485


Search completed: June  3, 2003, 14:08:24
Job time : 16.0184 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: June 3, 2003, 14:06:00 ; Search time 53.0696 Seconds
(without alignments)
1479.263 Million cell updates/sec

Title: US-09-804-690-2
Perfect score: 1996
Sequence: 1 MMSKYKYRDLTVRQTVNVIA.....FQLRALMQKARKTAGLSDLY 381

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1996	100.0	381	11	Q9CXS3	Q9cxs3 mus musculu
2	1894.5	94.9	390	4	Q9BUM5	Q9bum5 homo sapien
3	1813.5	90.9	392	13	Q9I8G8	Q9i8g8 chelonia my
4	964	48.3	408	5	Q9GSB5	Q9gsb5 drosophila
5	786	39.4	331	5	Q9VVA7	Q9vva7 drosophila
6	627.5	31.4	404	5	O76258	O76258 caenorhabdi
7	440	22.0	398	10	Q9LHG8	Q9lhg8 arabidopsis
8	412	20.6	174	11	O08761	O08761 mus musculu
9	389.5	19.5	368	10	Q9FFY6	Q9ffy6 arabidopsis
10	361	18.1	177	4	Q96FF5	Q96ff5 homo sapien
11	361	18.1	341	4	Q9NUX7	Q9nux7 homo sapien
12	293.5	14.7	385	3	P87279	P87279 saccharomyc
13	167.5	8.4	376	3	P78998	P78998 saccharomyc
14	151	7.6	148	4	Q9NTQ8	Q9ntq8 homo sapien
15	147	7.4	503	11	Q921F1	Q921f1 mus musculu
16	146	7.3	148	4	Q9NZ81	Q9nz81 homo sapien
17	143	7.2	746	4	Q9BU60	Q9bu60 homo sapien
18	143	7.2	1052	4	Q96FT1	Q96ft1 homo sapien
19	142.5	7.1	1021	4	O15451	O15451 homo sapien
20	142.5	7.1	1251	4	O15450	O15450 homo sapien
21	142	7.1	995	5	Q9NJ19	Q9nj19 aequipecten
22	141.5	7.1	485	3	Q9UTP6	Q9utp6 schizosacch
23	137.5	6.9	247	10	Q945K9	Q945k9 arabidopsis
24	137	6.9	397	4	Q96DL9	Q96dl9 homo sapien
25	135.5	6.8	463	11	Q922A2	Q922a2 mus musculu
26	135.5	6.8	827	3	O60130	O60130 schizosacch
27	135	6.8	137	11	Q9CQJ5	Q9cqj5 mus musculu
28	133	6.7	201	16	O86316	O86316 mycobacteri
29	132.5	6.6	472	3	O59907	O59907 neurospora
30	131.5	6.6	303	11	Q9D116	Q9d116 mus musculu
31	131.5	6.6	319	11	Q9D4W4	Q9d4w4 mus musculu
32	131.5	6.6	319	11	Q9D0D7	Q9d0d7 mus musculu
33	131	6.6	358	10	Q942Z3	Q942z3 oryza sativ
34	130.5	6.5	463	11	Q8VIN2	Q8vin2 rattus norv
35	130.5	6.5	1099	10	Q8W362	Q8w362 oryza sativ
36	130	6.5	609	17	Q8TXA4	Q8txa4 methanopyru
37	130	6.5	1006	11	Q62901	Q62901 rattus norv
38	129	6.5	866	5	Q8T8N7	Q8t8n7 caenorhabdi
39	129	6.5	924	5	Q9GYS7	Q9gys7 caenorhabdi
40	129	6.5	925	5	Q9BIC2	Q9bic2 caenorhabdi
41	128.5	6.4	303	11	Q8R3W0	Q8r3w0 mus musculu
42	127	6.4	279	10	O49490	O49490 arabidopsis
43	127	6.4	471	10	Q9LD31	Q91d31 crypthecodi
44	127	6.4	604	5	O97339	O97339 phascolion
45	126.5	6.3	331	10	Q93V84	Q93v84 arabidopsis

ALIGNMENTS

RESULT 1
Q9CXS3
ID Q9CXS3 PRELIMINARY; PRT; 381 AA.
AC Q9CXS3;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Tumor susceptibility gene 101.
GN TSG101.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYONIC HEAD;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK014049; BAB29131.1; -.
DR MGD; MGI:106581; tsg101.

```
DR   InterPro; IPR000608; UBQ_conjugat.
DR   SMART; SM00212; UBCc; 1.
SQ   SEQUENCE   381 AA;  43021 MW;  FA650A0BF7B21ABF CRC64;

  Query Match           100.0%;  Score 1996;  DB 11;  Length 381;
  Best Local Similarity  100.0%;  Pred. No. 3.2e-121;
  Matches  381;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNI 60

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQI 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQI 120

Qy        121 MIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCP 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 MIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCP 180

Qy        181 YPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 YPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGA 240

Qy        241 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 300

Qy        301 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 360

Qy        361 QFQLRALMQKARKTAGLSDLY 381
              |||||||||||||||||||||
Db        361 QFQLRALMQKARKTAGLSDLY 381


RESULT 2
Q9BUM5
ID   Q9BUM5      PRELIMINARY;      PRT;   390 AA.
AC   Q9BUM5;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Tumor susceptibility gene 101.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EYE;
RA   Strausberg R.;
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC002487; AAH02487.1; -.
DR   InterPro; IPR000608; UBQ_conjugat.
DR   SMART; SM00212; UBCc; 1.
SQ   SEQUENCE   390 AA;  43910 MW;  96BB2A4FC22DF16A CRC64;

  Query Match           94.9%;  Score 1894.5;  DB 4;  Length 390;
  Best Local Similarity  94.2%;  Pred. No. 1.2e-114;
  Matches  359;  Conservative   11;  Mismatches   10;  Indels    1;  Gaps    1;

Qy          1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNI 60
              |:||||||||||:|||| :|||||||||||||||||:|||||||| |||| |||
Db         11 MVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSRELMNLTGTIPVPYRGNTYNI 70

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQI 120
              ||||||||||||||||||||||||||||||||||:||||:|:|| |||:
Db         71 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLLGLIQV 130

Qy        121 MIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCP 180
              |||:||:|||||||| :||||||| |||||||||||||| ||| ||||||||||||||||
Db        131 MIVVFGDEPPVFSRP-ISASYPPYQATGPPNTSYMPGMPGGISPYPSGYPPNPSGYPGCP 189

Qy        181 YPPAGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGA 240
              ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        190 YPPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDRA 249

Qy        241 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        250 QAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQS 309

Qy        301 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRK 360
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        310 ENNDIDEVIIPTAPLYKQILNLYAEENAIEDTILYLGEALRRGVIDLDVFLKHVRLLSRK 369

Qy        361 QFQLRALMQKARKTAGLSDLY 381
              |||||||||||||||||||||
Db        370 QFQLRALMQKARKTAGLSDLY 390


RESULT 3
Q9I8G8
ID   Q9I8G8      PRELIMINARY;      PRT;   392 AA.
AC   Q9I8G8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Tumor susceptibility protein 101.
GN   TSG101.
OS   Chelonia mydas caranigra (Green sea-turtle).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Testudines; Cryptodira; Chelonioidea; Cheloniidae; Chelonia.
OX   NCBI_TaxID=8469;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Yu Q., Lu Y., Nerurkar V.R., Yanagihara R.;
RT   "Studies on the turtle tumor susceptibility gene, TSG; full-length
RT   cDNA sequence, genomic structure analysis and role in green turtle
RT   fibropapilloma.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF279276; AAF87776.1; -.
DR   InterPro; IPR000608; UBQ_conjugat.
DR   SMART; SM00212; UBCc; 1.
SQ   SEQUENCE   392 AA;  44314 MW;  25DE98A5116CF8EA CRC64;

  Query Match           90.9%;  Score 1813.5;  DB 13;  Length 392;
  Best Local Similarity  89.8%;  Pred. No. 2e-109;
  Matches  343;  Conservative   20;  Mismatches   18;  Indels    1;  Gaps    1;

Qy          1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNI 60
              |::|||||||||::| :||  |||||||:|:|||||||||:|::||||||| |||| |||
Db         11 MLAKYKYRDLTVQETTSVITQYKDLKPVMDAYVFNDGSSRDLMSLTGTIPVPYRGNTYNI 70

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQI 120
              ||||||||||:|||||||||||||||||||||||||||||||:||||:|:|: ||||
Db         71 PICLWLLDTYPFNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHEWKHPQSDLIGLIQI 130

Qy        121 MIVIFGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCP 180
              |||:||||||||||:| |: || |||||||||||||||||| || |:|||||||| |
Db        131 MIVVFGEEPPVFSRPTISTSFQPYQATGPPNTSYMPGMPSGISPYPPGHPPNPSGYPGYP 190

Qy        181 YPPAGPYPATTSSQ-YPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDG 239
              ||| ||:||||| | | |||||||||||||||||||||||| ||||||||||||||||
Db        191 YPPGGPFPATTSGQHYTSQPPVTTVGPSRDGTISEDTIRASRISAVSDKLRWRMKEEMDR 250

Qy        240 AQAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQ 299
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        251 AQAELNALKRTEEDLKKGHQKLEEMVTRLDHEVAEVDKNIELLKKKDEELSSALEKMENQ 310

Qy        300 SENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSR 359
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        311 SENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSR 370

Qy        360 KQFQLRALMQKARKTAGLSDLY 381
```

```
                    ||||||||||||||||||||||
Db        371 KQFQLRALMQKARKTAGLSDLY 392


RESULT 4
Q9GSB5
ID   Q9GSB5      PRELIMINARY;      PRT;   408 AA.
AC   Q9GSB5;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Tumor suppressor protein 101.
GN   TSG101 OR CG9712.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Laurencon A., Burtis K.C., Hawley S.;
RT   "TSG101 fly homolog.";
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF315343; AAG29564.1; -.
DR   FlyBase; FBgn0036666; TSG101.
DR   InterPro; IPR001440; TPR.
DR   InterPro; IPR000608; UBQ_conjugat.
DR   SMART; SM00212; UBCc; 1.
SQ   SEQUENCE   408 AA;  45190 MW;  7C48E20DEBB29A19 CRC64;

  Query Match             48.3%;  Score 964;  DB 5;  Length 408;
  Best Local Similarity   49.3%;  Pred. No. 1.5e-54;
  Matches  199;  Conservative   66;  Mismatches  103;  Indels   36;  Gaps    9;

Qy          2 MSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNIP 61
              :|||||   | :  |:|:  :: | :  |  :|||||||:||  : ||||| |: | | ||
Db         13 LSKYKYVAATKKDVVDVVTSFRSLTYDLQRFVFNDGSSKELFTIQGTIPVVYKNNTYYIP 72

Qy         62 ICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQIM 121
              ||:||:||:| | |:|||||| :| ||   :|| |||:|||||||||:   |:|| |||:|
Db         73 ICIWLMDTHPQNAPMCFVKPTPTMQIKVSMYVDHNGKVYLPYLHDWQPHSSDLLSLIQVM 132

Qy        122 IVIFGEEPPVFSRP--TVSASYPPYTATGPPNTSYM--PGMPSGISA---YP-------S 167
              || ||: |||:|:|   ::| |        |  |||  || | | ::   ||       |
Db        133 IVTFGDHPPVYSKPKEQIAAPY--------PTNSYMPQPGAPGGSNSFLPYPTAGGAGGS 184

Qy        168 GYPPNPSGYPGCPYP--PAGP-----YPA------TTSSQYPSQPPVTTVGPSRDGTISE 214
              :|| |:|    |||  |||| : |||       |:  ||          ||  |||:|
Db        185 NFPPYPTGSNVGPYPPTPAGPAGGSGYPAYPNFIQPTAGGYPPAAGYNPSNPSSTGTITE 244

Qy        215 DTIRASLISAVSDKLRWRMKEEMDGAQAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAE 274
              : |:||:|||: |||| |::|:::  |||:  | ||:::| :|  |:: :: ||::|  :
Db        245 EHIKASIISAIDDKLRRRVQEKVNQYQAEIETLNRTKQELLEGSAKIDAIIERLEREHID 304

Qy        275 VDKNIELLKKKDEELSSALEKMENQSENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIF 334
              : ||| :|| |::||  ||| :|:    |  || :  |||||:|:|| ||:| | || |:
Db        305 MQKNISILKDKEQELEKALEDLESAEAINP-DEAVTTTAPLYRQLLNAYADEAATEDAIY 363

Qy        335 YLGEALRRGVIDLDVFLKHVRLLSRKQFQLRALMQKARKTAGLS 378
              |||| || ||||| |||||| |||||| ||| ||| |: |||:
Db        364 YLGEGLRGGVIDLETFLKHVRQLSRKQFILRATMQKCRQKAGLA 407


RESULT 5
Q9VVA7
ID   Q9VVA7      PRELIMINARY;      PRT;   331 AA.
AC   Q9VVA7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CG9712 protein.
GN   TSG101 OR CG9712.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003526; AAF49406.1; -.
DR   FlyBase; FBgn0036666; TSG101.
DR   InterPro; IPR001440; TPR.
SQ   SEQUENCE   331 AA;  36255 MW;  A6DDDA84C718B738 CRC64;

  Query Match             39.4%;  Score 786;  DB 5;  Length 331;
  Best Local Similarity   49.0%;  Pred. No. 3.6e-43;
  Matches  166;  Conservative   55;  Mismatches   82;  Indels   36;  Gaps    9;

Qy         67 LDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQIMIVIFG 126
              :||:| | |:|||||| :| ||   :|| |||:||||||||:   |:|| |||:||| ||
Db          1 MDTHPQNAPMCFVKPTPTMQIKVSMYVDHNGKVYLPYLHDWQPHSSDLLSLIQVMIVTFG 60

Qy        127 EEPPVFSRP--TVSASYPPYTATGPPNTSYM--PGMPSGISA---YP-------SGYPPN 172
              : |||:|:|   ::| |        |  |||  || | | ::   ||       | :||
Db         61 DHPPVYSKPKEQIAAPY--------PTNSYMPQPGAPGGSNSFLPYPTAGGAGGSNFPPY 112

Qy        173 PSGYPGCPYP--PAGP-----YPA------TTSSQYPSQPPVTTVGPSRDGTISEDTIRA 219
              |:|    |||  ||||     |||       |:  ||          ||  |||:|: |:|
Db        113 PTGSNVGPYPPTPAGPAGGSGYPAYPNFIQPTAGGYPPAAGYNPSNPSSTGTITEEHIKA 172

Qy        220 SLISAVSDKLRWRMKEEMDGAQAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNI 279
              |:|||: |||| |::|:::  |||:  | ||:::| :|  |:: :: ||::|  :: |||
Db        173 SIISAIDDKLRRRVQEKVNQYQAEIETLNRTKQELLEGSAKIDAIIERLEREHIDMQKNI 232

Qy        280 ELLKKKDEELSSALEKMENQSENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEA 339
               :|| |::||  ||| :|:    |  || :  |||||:|:|| ||:| | || |:||||
```

```
Db        233 SILKDKEQELEKALEDLESAEAINP-DEAVTTTAPLYRQLLNAYADEAATEDAIYYLGEG 291

Qy        340 LRRGVIDLDVFLKHVRLLSRKQFQLRALMQKARKTAGLS 378
              || |||||: |||||| |||||| ||| ||| |: |||:
Db        292 LRGGVIDLETFLKHVRQLSRKQFILRATMQKCRQKAGLA 330


RESULT 6
O76258
ID   O76258      PRELIMINARY;      PRT;   404 AA.
AC   O76258;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   C09G12.9 protein.
GN   C09G12.9.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Le T.T., Harper M.;
RT   "The sequence of C. elegans cosmid C09G12.";
RL   Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF038608; AAC25822.1; -.
DR   InterPro; IPR000608; UBQ_conjugat.
DR   SMART; SM00212; UBCc; 1.
SQ   SEQUENCE   404 AA;  43909 MW;  2C322EC52DF4C62D CRC64;

  Query Match           31.4%;  Score 627.5;  DB 5;  Length 404;
  Best Local Similarity  36.1%;  Pred. No. 7.8e-33;
  Matches  145;  Conservative   78;  Mismatches  134;  Indels   45;  Gaps   10;

Qy          6 KYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNIPICLW 65
              || |   :  :  :: :||| |  |:::| ||  |    | ||||| |:|  |||| :: 
Db         16 KYADSAKKDIIGALSQFKDLSPGTDTFMFPDGKRRTAFRLKGTIPVYYKGACYNIPVTVY 75

Qy         66 LLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQIMIVIF 125
              | ||:||  |||:| |||:|     :||:  ||::|||||::|: |  :|  |:|::
Db         76 LWDTHPYYAPICYVNPTSTM---ESEHVNKEGKVFLPYLNEWRFPGYDLSGLLQMIFARS 132

Qy        126 GEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSG-----YPPNPSGYPGC- 179
                  :  |  ::    ::|  |  |  | ||   : ||:|     |||: : ||
Db        133 AANSATNASATNPSAGSSASSTPTPYPSSQPTMP---TPYPTGSGAAPYPPSSTPYPSAG 189

Qy        180 -----PY---PPAGPYP--ATTSSQYPS----------QPPV---TTVGPSRDGTISED 215
                   ||   | : |||  |: :| |||           |||   |:|  |  |||  |
Db        190 AMGYNPYMNVPQSTPYPMGASGASPYPSASSNPAPPPPRPPPVTAQTSVSSSSGGTIQAD 249

Qy        216 TIRASLISAVSDKLRWRMKEEMDGAQAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEV 275
              |||||::||| :|:| :::| |    ||: ::: | ::|::| |||: |:  |: : : :
Db        250 TIRASVMSAVEEKIRAKLRERMGTNSAEMASIRTTSDELREGQQKLKRMLEELETQRSSL 309

Qy        276 DKNIELLKKKDEELSSALEKMENQSENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFY 335
                :   |  ||: ||      ::   ||| |    ||::||:  ||::
Db        310 QTACEIYTAKKAELAKALSD-AGGTDAPPIDEAIDAAFPLHRQIVLNYAKDLTC------ 362

Qy        336 LGEALRRGVIDLDVFLKHVRLLSRKQFQLRALMQKARKTAGL 377


                ::|::  | |  :|:||| :||:||  || ||| |:||||
Db        363 --QSLKKRQITLAEYLRHVRDVSREQFIYRATMQKCRRTAGL 402


RESULT 7
Q9LHG8
ID   Q9LHG8      PRELIMINARY;      PRT;   398 AA.
AC   Q9LHG8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Mouse and human tumor susceptibility gene-like protein (Hypothetical
DE   44.7 kDa protein).
GN   T2E22.28 OR T2E22_129/AT3G12400.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RA   Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RX   PubMed=10907853;
RA   Nakamura Y.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT   Sequence features of the regions of 4,251,695 bp covered by ninety P1,
RT   TAC and BAC clones.";
RL   DNA Res. 7:217-221(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE=21016720; PubMed=11130713;
RA   Salanoubat M., Lemcke K., Rieger M., Ansorge W., Unseld M.,
RA   Fartmann B., Valle G., Bloecker H., Perez-Alonso M., Obermaier B.,
RA   Delseny M., Boutry M., Grivell L.A., Mache R., Puigdomenech P.,
RA   De Simone V., Choisne N., Artiguenave F., Robert C., Brottier P.,
RA   Wincker P., Cattolico L., Weissenbach J., Saurin W., Quetier F.,
RA   Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
RA   Wurmbach E., Drzonek H., Erfle H., Jordan N., Bangert S.,
RA   Wiedelmann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
RA   Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simionati B.,
RA   Conrad A., Hornischer K., Kauer G., Loehnert T.-H., Nordsiek G.,
RA   Reichelt J., Scharfe M., Schoen O., Bargues M., Terol J., Climent J.,
RA   Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,
RA   Cooke R., Laudie M., Berger-Llauro C., Purnelle B., Masuy D.,
RA   de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
RA   Monfort A., Argiriou A., Flores M., Liguori R., Vitale D.,
RA   Mannhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,
RA   Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
RA   Rooney T., Rizzo M., Walts A., Utterback T., Fujii C.Y., Shea T.P.,
RA   Creasy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
RA   Pai G., Militscher J., Sellers P., Gill J.E., Feldblyum T.V.,
RA   Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
RA   Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
RA   Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
RA   Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA   Nakayama S., Nakazaki N., Shinpo S., Takeuchi C., Wada T.,
RA   Watanabe A., Yamada M., Yasuda M., Tabata S.;
RT   "Sequence and analysis of chromosome 3 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 408:820-822(2000).
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Yamada K., Banh J., Banno F., Dale J.M., Goldsmith A.D., Lee J.M.,
RA   Onodera C.S., Quach H.L., Tang C., Toriumi M., Yamamura Y., Yu G.,
RA   Yu S., Bowser L., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,
RA   Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Kim C.,
RA   Koesema E., Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M.,
```

```
RA   Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shinn P.,
RA   Southwick A., Tracy S.E., Shinozaki K., Davis R.W., Ecker J.R.,
RA   Theologis A.;
RT   "Full Length cDNA of gene T2E22_129/At3g12400 (GI:15230478).";
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP002047; BAB03147.1; -.
DR   EMBL; AC069474; AAG51025.1; -.
DR   EMBL; AY056283; AAL07132.1; -.
DR   InterPro; IPR000608; UBQ_conjugat.
DR   SMART; SM00212; UBCc; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   398 AA;  44715 MW;  659996F9784FDC20 CRC64;

  Query Match             22.0%;  Score 440;  DB 10;  Length 398;
  Best Local Similarity   31.6%;  Pred. No. 9.8e-21;
  Matches  117;  Conservative   72;  Mismatches  147;  Indels   34;  Gaps   12;

Qy         12 VRQ-TVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNIPICLWLLDTY 70
              :|| :|:|: |  |:|   |:: ||| |  |:   ||||: : |  |||||: :|||::|
Db         37 IRQHLLNLISSYPSLEPKTASFMHNDGRSVNLLQADGTIPMPFHGVTYNIPVIIWLLESY 96

Qy         71 PYNPPICFVKPTSSMTIK-TGKHVDANGKIYLPYLHDWKHPRSELLELIQIMIVIFGEEP 129
              | :||  :| ||: | ||    ||  :| : |||| :| :| | |::|:  :   |  :|
Db         97 PRHPPCVYVNPTADMIIKRPHAHVTPSGLVSLPYLQNWVYPSSNLVDLVSDLSAAFARDP 156

Qy        130 PVFSRPTVSASYPPYTATGPPNT--SYMPGMPSGISAYPSGYPPNPSGYPGCPYPPAGPY 187
              |::||       ||     ||    | :   ||   : |  :||:| |          |
Db        157 PLYSR---RRPQPP--PPSPPTVYDSSLSRPPSADQSLPRPFPPSPYG--------GGVS 203

Qy        188 PATTSSQYPSQPPVTTVGPSRDGTISE--DTIRASLISAVSDKLRWRMKEEMDGAQAELN 245
                 : |          :   |::  : : : |:|    : |    ||:   ||
Db        204 RVQVQHVHHQQQSDDAAEVFKRNAINKMVEMVHSDLVS--MRRAREAEAEELLSLQA--- 258

Qy        246 ALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKMENQSEN--- 302
              ||| |::|  |   |:|||    :|   ::: ::::     : | : : : :::|
Db        259 GLKRREDELNIG---LKEMV----EEKETLEQQLQIISMNTDILDSWVRENQGKTKNLVD 311

Qy        303 NDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRKQF 362
              |:|       | ||:|   | : |||| |: | :: : ||:  | :|::|||||||:||
Db        312 LDVDNAFECGDTLSKQMLECTALDLAIEDAIYSLDKSFQDGVVPFDQYLRNVRLLSREQF 371

Qy        363 QLRALMQKAR 372
                ||   | |
Db        372 FHRATGSKVR 381


RESULT 8
O08761
ID   O08761      PRELIMINARY;      PRT;   174 AA.
AC   O08761;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Signaling molecule.
GN   ATTP OR ATTP.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=THYMUS;
RA   Burbelo P.D.;
RT   "Role of ATTP in signal transduction.";
RL   Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U97571; AAB57699.1; -.
DR   MGD; MGI:1860490; Attp.
DR   InterPro; IPR000608; UBQ_conjugat.
DR   SMART; SM00212; UBCc; 1.
SQ   SEQUENCE   174 AA;  20499 MW;  81BEE363DD81900B CRC64;

  Query Match             20.6%;  Score 412;  DB 11;  Length 174;
  Best Local Similarity   53.3%;  Pred. No. 2.3e-19;
  Matches   72;  Conservative   27;  Mismatches   36;  Indels    0;  Gaps    0;

Qy          1 MMSKYKYRDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNI 60
              :: |||:|||||  :  ||   :   :|:||| | | ::|:| ||||||  |:|  |||
Db         11 LLGKYKFRDLTVEELKNVSVSFPHFRYSVDTYVFKDTSQKDLLNFTGTIPVMYQGKTYNI 70

Qy         61 PICLWLLDTYPYNPPICFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQI 120
              ||  |:||::|: |||||:|||::| |  |||||| |:|||||| :| ||:| :: ||:
Db         71 PIRFWILDSHPFAPPICFLKPTANMEISVGKHVDAKGRIYLPYLQNWSHPKSAIVGLIKE 130

Qy        121 MIVIFGEEPPVFSRP 135
              ||  | || |::| |
Db        131 MIARFQEELPLYSIP 145


RESULT 9
Q9FFY6
ID   Q9FFY6      PRELIMINARY;      PRT;   368 AA.
AC   Q9FFY6;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Similarity to nascent polypeptide associated complex alpha chain.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RX   MEDLINE=97471969; PubMed=9330910;
RA   Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,
RA   Miyajima N., Tabata S.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence
RT   features of the 1.6 Mb regions covered by twenty physically assigned
RT   P1 clones.";
RL   DNA Res. 4:215-230(1997).
DR   EMBL; AB005230; BAB11114.1; -.
DR   InterPro; IPR000608; UBQ_conjugat.
DR   SMART; SM00212; UBCc; 1.
SQ   SEQUENCE   368 AA;  41495 MW;  262F8CDB45BBCD40 CRC64;

  Query Match             19.5%;  Score 389.5;  DB 10;  Length 368;
  Best Local Similarity   28.3%;  Pred. No. 1.6e-17;
  Matches  108;  Conservative   71;  Mismatches  124;  Indels   79;  Gaps   12;

Qy          8 RDLTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNIPICLWLL 67
              : |  :  :|:|: |  | |   ::  ||| |  |:   |||| : ::|  ||||: :|||
Db         34 KSLIRQHLLNLISSYTSLDPKTATFTHNDGRSVILLQADGTIPMPFQGVSYNIPVVIWLL 93

Qy         68 DTYPYNPPICFVKPTSSMTIK-TGKHVDANGKIYLPYLHDWKHPRSELLELIQIMIVIFG 126
              ::||  ||  :| ||  | ||     :|  :| : |||| :| :| | |::|   :   |
Db         94 ESYPQYPPCVYVNPTRDMIIKRPHSNVSPSGLVSLPYLQNWIYPSSNLVDLASHLSAAFS 153

Qy        127 EEPPVFS--RPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCPYPPA 184
               :||::|  ||             ||  |  | : || |                 | ||
Db        154 RDPPLYSQRRP-------------PPQPS--PSIGSGYSR--------------PLPP- 182

Qy        185 GPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDK-------LRWRMKEEM 237
                                      |    :  :  : |: : :         :|   : |
Db        183 ----------------------RQTDDAAEVYKKNAINRIVEMVHGDIVLMRSAREVET 219

Qy        238 DGAQAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELSSALEKME 297
              :|  :  : ||| ||:: |    :|||  :::|  |  : ::::    : | | :   |
Db        220 EGLLSLQSDLKRREEEINNG---FKEMV--IEKETLE--QQLQVIAMNTDVLGSWIR--E 270

Qy        298 NQSE------NNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFL 351
              || :      : |:|:      | ||:|   | : |||| :: : :: | | :  | :|
Db        271 NQGKAKDLLVDLDVDDSFECIDSLSKQMLECTALDLAIEDVVYSMDKSFRDGSLPFDQYL 330

Qy        352 KHVRLLSRKQFQLRALMQKARK 373
              ::|||||:||  ||  :| |:
```

```
Db        331 RNVRLLSREQFFHRATAEKVRE 352


RESULT 10
Q96FF5
ID   Q96FF5      PRELIMINARY;      PRT;   177 AA.
AC   Q96FF5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Unknown (protein for MGC:13394).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RA   Strausberg R.;
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC011011; AAH11011.1; -.
SQ   SEQUENCE   177 AA;  19552 MW;  E733E710F8D0068A CRC64;

  Query Match             18.1%;  Score 361;  DB 4;  Length 177;
  Best Local Similarity   59.0%;  Pred. No. 4.6e-16;
  Matches   62;  Conservative   21;  Mismatches   22;  Indels    0;  Gaps    0;

Qy         29 LDSYVFNDGSSRELVNLTGTIPVRYRGNIYNIPICLWLLDTYPYNPPICFVKPTSSMTIK 88
              :|:||| | | ::|:| |||||| |:|| |||||  |:||::|: |||||:|||::| |
Db          1 MDTYVFKDSSQKDLLNFTGTIPVMYQGNTYNIPIRFWILDSHPFAPPICFLKPTANMGIL 60

Qy         89 TGKHVDANGKIYLPYLHDWKHPRSELLELIQIMIVIFGEEPPVFS 133
              |||||| |:|||||| :| ||:| :: ||: ||  | || |::|
Db         61 VGKHVDAQGRIYLPYLQNWSHPKSVIVGLIKEMIAKFQEELPMYS 105


RESULT 11
Q9NUX7
ID   Q9NUX7      PRELIMINARY;      PRT;   341 AA.
AC   Q9NUX7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   CDNA FLJ11068 fis, clone PLACE1004918, weakly similar to L-lactate
DE   dehydrogenase M chain (EC 1.1.1.27).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RA   Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA   Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA   Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA   Nakamura Y., Nagahari K., Masuho Y., Sasaki N.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK001930; BAA91985.1; -.
DR   HSSP; P00336; 5LDH.
DR   InterPro; IPR001236; ldh.
DR   InterPro; IPR001557; L_LDH.
DR   InterPro; IPR000205; NAD_binding.
DR   InterPro; IPR000608; UBQ_conjugat.
DR   Pfam; PF00056; ldh; 1.
DR   Pfam; PF02866; ldh_C; 1.
DR   PRINTS; PR00086; LLDHDRGNASE.
DR   SMART; SM00212; UBCc; 1.
SQ   SEQUENCE   341 AA;  37587 MW;  7D5CD535296165FA CRC64;

  Query Match             18.1%;  Score 361;  DB 4;  Length 341;
  Best Local Similarity   59.0%;  Pred. No. 1e-15;
  Matches   62;  Conservative   21;  Mismatches   22;  Indels    0;  Gaps    0;

Qy         29 LDSYVFNDGSSRELVNLTGTIPVRYRGNIYNIPICLWLLDTYPYNPPICFVKPTSSMTIK 88
              :|:||| | | ::|:| |||||| |:|| |||||  |:||::|: |||||:|||::| |
Db          1 MDTYVFKDSSQKDLLNFTGTIPVMYQGNTYNIPIRFWILDSHPFAPPICFLKPTANMGIL 60

Qy         89 TGKHVDANGKIYLPYLHDWKHPRSELLELIQIMIVIFGEEPPVFS 133
              |||||| |:|||||| :| ||:| :: ||: ||  | || |::|
Db         61 VGKHVDAQGRIYLPYLQNWSHPKSVIVGLIKEMIAKFQEELPMYS 105


RESULT 12
P87279
ID   P87279      PRELIMINARY;      PRT;   385 AA.
AC   P87279;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   STP22P.
GN   STP22 OR YCL008C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=XG1#1;
RA   Li Y., Kane T., Tipper C., Spatrick P., Jenness D.D.;
RT   "The complete sequence of STP22 gene.";
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF004731; AAB62820.1; -.
DR   SGD; S0000514; STP22.
DR   InterPro; IPR000608; UBQ_conjugat.
DR   SMART; SM00212; UBCc; 1.
DR   PROSITE; PS50127; UBIQUITIN_CONJUGAT_2; 1.
SQ   SEQUENCE   385 AA;  43330 MW;  FEDE3BE79F063BCE CRC64;

  Query Match             14.7%;  Score 293.5;  DB 3;  Length 385;
  Best Local Similarity   26.0%;  Pred. No. 2.7e-11;
  Matches  102;  Conservative   74;  Mismatches  164;  Indels   53;  Gaps   16;

Qy          7 YRD--LTVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNI-YNIPIC 63
              | |   |   :: ::  :  |:|    :  :||: : |::: |||     |:  ::||:
Db         25 YNDGRTTFHDSLALLDNFHSLRPRTRVFTHSDGTPQLLLSIYGTISTGEDGSSPHSIPVI 84

Qy         64 LWLLDTYPYNPPIC------FVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLEL 117
              :|:   ||  ||        |   | | ::   :::|:|| | || || |      |: :
Db         85 MWVPSMYPVKPPFISINLENFDMNTISSSLPIQEYIDSNGWIALPILHCWDPAAMNLIMV 144

Qy        118 IQIMIVIFGEEPPVFSRPTVSASYPPYTAT---GPPNTSYMPGMPSGISAYPSGYPPNPS 174
              :| :: :  |||   |   | ||   |      ||  :|  |      |   || |
Db        145 VQELMSLL-HEPPQDQAP----SLPPKPNTQLQQEQNTPPLPPKPKS----PHLKPPLP- 194

Qy        175 GYPGCPYPPAGPYPATTSSQYPSQPPVTTVGPSRD-------GTISEDTIRASLISAVSD 227
                    ||  | ||: :         | : |:          |:  :  |   :  |:|
Db        195 -------PPPPPQPASNALDLMDMDN-TDISPTNHHEMLQNLQTVVNELYRED-VDYVAD 245

Qy        228 KL--RWRMKEEMDGAQAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKK 285
              |:  |  : :|      |: |: :    |:   | :|: :  |: ::  :  |    : |
Db        246 KILTRQTVMQESIARFHEIIAIDKNH--LRAVEQAIEQTMHSLNAQIDVLTAN----RAK 299

Qy        286 DEELSSALEKMENQSENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVI 345
              :: ||      :  :: |:: : :       |: || |:: |: ||| |   | || |
Db        300 VQQFSST-----SHVDDEDVNSIAVAKTDGLNQLYNLVAQDYALTDTIECLSRMLHRGTI 354

Qy        346 DLDVFLKHVRLLSRKQFQLRALMQKARKTAGLS 378
              || |:|  | |:|:|| :|  :|  | |: ||
Db        355 PLDTFVKQGRELARQQFLVRWHIQ--RITSPLS 385


RESULT 13
P78998
ID   P78998      PRELIMINARY;      PRT;   376 AA.
AC   P78998;
```

```
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Hypothetical 41.5 kDa protein.
GN   LGYCLUN5C.
OS   Saccharomyces pastorianus (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=27292;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Andersen T., Nilsson-Tillgren T.;
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z86109; CAB06793.1; -.
DR   InterPro; IPR001284; Ribosomal_L34E.
DR   InterPro; IPR000608; UBQ_conjugat.
DR   SMART; SM00212; UBCc; 1.
DR   PROSITE; PS01145; RIBOSOMAL_L34E; UNKNOWN_1.
DR   PROSITE; PS50127; UBIQUITIN_CONJUGAT_2; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   376 AA;  41542 MW;  837F91DA97BF0A0B CRC64;

  Query Match             8.4%;  Score 167.5;  DB 3;  Length 376;
  Best Local Similarity   19.3%;  Pred. No. 0.0036;
  Matches   64;  Conservative   59;  Mismatches  127;  Indels   81;  Gaps    9;

Qy         11 TVRQTVNVIAMYKDLKPVLDSYVFNDGSSRELVNLTGTIPVRYRGNIYNIPICLWLLDTY 70
              |   :: ::  :  |:|    :  :|||  : |::: ||:     |:  ::|: :|:   |
Db         31 TFHDSLALLDKFHRLRPRTRVFTHSDGSPQLLLSIYGTV-----GD--SLPLIMWIPSLY 83

Qy         71 PYNPPI------CFVKPTSSMTIKTGKHVDANGKIYLPYLHDWKHPRSELLELIQIMIVI 124
              |  ||       |     | ::    ::|::|:: || |  |      |: ::| :: :
Db         84 PVKPPFISIDLETFDVSAISSSLPVQAYIDSDGRVALPILDHWNPATMGLIVVVQELMSL 143

Qy        125 FGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCPYPPA 184
              |                    |:   :| :|    |  |     |  :   | ||
Db        144 LSE--------------------PSRDRVPSLPPKPDAMHS-----PLKHTVPPLPPK 176

Qy        185 GPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAV-------SDKLRWRMKEEM 237
              :|     | | |||     ||               |||       :|       | :
Db        177 AKFPHV--QPPLQPPPPPQPPS--------------SAVDLMDMDNTDLSPTNHHEML 218

Qy        238 DGAQAELNALKR---------------TEEDLKKGHQKLEEMVTRLDQEVAEVDKNIEL 281
                |: || | |                 :: : : |: :    |||      ::: :
Db        219 QNLQSVLNELYREDVHYVADKILTRQTIMQDSVARFHEMVAVDKTRLQAVEQTIEQTMHT 278

Qy        282 LKKKDEELSSALEKMENQSENNDIDEVIIPT 312
              |  : | |::    ::  |  :  |:  : |
Db        279 LNAQIEVLTAERATVQEFSSTSPFDDEDVDT 309


RESULT 14
Q9NTQ8
ID   Q9NTQ8       PRELIMINARY;      PRT;   148 AA.
AC   Q9NTQ8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Hypothetical 15.4 kDa protein.
GN   DKFZP564J157.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=21154917; PubMed=11230166;
RA   Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,
RA   Ansorge W., Boecher M., Bloecker H., Bauersachs S., Blum H.,
RA   Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,
RA   Mewes H.W., Ottenwaelder B., Obermaier B., Tampe J., Heubner D.,
RA   Wambutt R., Korn B., Klein M., Poustka A.;
RT   "Towards a Catalog of Human Genes and Proteins: Sequencing and
RT   Analysis of 500 Novel Complete Protein Coding Human cDNAs.";
RL   Genome Res. 11:422-435(2001).
DR   EMBL; AL122042; CAB59179.2; -.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR01217; PRICHEXTENSN.
KW   Hypothetical protein.
SQ   SEQUENCE   148 AA;  15375 MW;  796D11BA16A0591E CRC64;

  Query Match             7.6%;  Score 151;  DB 4;  Length 148;
  Best Local Similarity   28.7%;  Pred. No. 0.013;
  Matches   39;  Conservative   15;  Mismatches   32;  Indels   50;  Gaps    6;

Qy        135 PTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNP---SGYPGC----PYPPAGPY 187
              | ::  :||      |       || | |  |:||| ||:|    |||||    ||||  |
Db         30 PPINPPFPPGPCPPP------PGAPHGNPAFPSGGPPHPVPQPGYPGCQPLGPYPPPYPP 83

Qy        188 PATTSSQYPSQPPVTTVGPSRDGTISEDTIRASLISAVSDKLRWRMKEEMDGAQAELNAL 247
              ||      |  |||  : |   |             : : |
Db         84 PA------PGIPPVNPLAPGMVGP----------AVIVD------------------- 106

Qy        248 KRTEEDLKKGHQKLEE 263
              |: :: :|| |:|: :
Db        107 KKMQKKMKKAHKKMHK 122


RESULT 15
Q921F1
ID   Q921F1       PRELIMINARY;      PRT;   503 AA.
AC   Q921F1;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Similar to annexin A11.
GN   ANXA11.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC012875; AAH12875.1; -.
DR   MGD; MGI:108481; Anxa11.
DR   InterPro; IPR001464; Annexin.
DR   Pfam; PF00191; annexin; 4.
DR   PRINTS; PR00196; ANNEXIN.
DR   ProDom; PD000143; Annexin; 4.
DR   PROSITE; PS00223; ANNEXIN; UNKNOWN_4.
SQ   SEQUENCE   503 AA;  54079 MW;  33F3471EE21A0D32 CRC64;

  Query Match             7.4%;  Score 147;  DB 11;  Length 503;
  Best Local Similarity   26.3%;  Pred. No. 0.11;
  Matches   77;  Conservative   29;  Mismatches   95;  Indels   92;  Gaps   16;

Qy        125 FGEEPPVFSRPTVSASYPPYTATGPPNTSYMPGMPSGISAYPSGYPPNPSGYPGCPYPPA 184
              ||: ||   :|      |||    ||  :  |||||    ||: ||  |   || | ||
Db         89 FGQPPPA-QQPV-----PPYGMYPPPGGNPPPGMPS----YPA-YPGAP--VPGQPMPPT 135

Qy        185 GPYPATTSSQYPSQPPVTTVG-----------PSRDGTISEDTIRASL----------I 222
              |  |       || |||:|  |           ||  |     ||  ::          |
Db        136 GQQP---PGAYPGQPPMTYPGQSPMPPPGQQPVPSYPGYSGSSTITPAVPPAQFGNRGTI 192

Qy        223 SAVS--------DKLRWRMKEEMDGAQAELNALKRTEEDLKKGHQKLEEMVTRLDQEVAE 274
              :| |        : ||  ||       || :: |       |:  | |    |   :
Db        193 TAASGFDPLRDAEVLRKAMKGFGTDEQAIIDCLGSRSN--KQRQQILLSFKTAYGK---- 246

Qy        275 VDKNIELLKKKDEELSSALEKMENQSENNDIDEVIIPTAPLYKQILNLYAEENAIEDTIF 334
                :|:|    |||   |                        | || |   :  :   ::
Db        247 -----DLIKDLKSELSGNFE--------------------KTILALM--KTPVLFDVY 277

Qy        335 YLGEALRRGVID----LDVFL----KHVRLLSRKQFQLRALMQKARKTAGLSD 379
```

```
                : ||::    |    :::|     :|:| |||     :   ||  : |  ||
Db        278 EIKEAIKGAGTDEACLIEIFASRSNEHIRELSR---AYKTEFQKTLEEAIRSD 327
```

Search completed: June 3, 2003, 14:11:50
Job time : 56.0696 secs